From Micro to Macro: Using a Granular Lens to Analyze Global Complex Systems

Valentina Semenova[1,2,3]

[1]*Department of Mathematics, University of Oxford*
[2]*Institute of New Economic Thinking at the Oxford Martin School*
[3]*Oxford-Man Institute of Quantitative Finance*

Abstract

In this thesis, I take a complex systems perspective on the economy with the aim of tying granular activity and social behaviours among individuals to global economic and political phenomena. The thesis leverages tools from network science, machine learning, and dynamical systems to extract individual activity from noisy real-world data and subsequently model the macro-scale consequences of individual decisions.

Several chapters of the thesis focus on social media activity and the financial system. I focus on the online investor discussion forum called `WallStreetBets` (WSB). The thesis first presents a data-driven analysis of WSB: initial experiments decompose the forum using a large language topic model and network tools. The findings show that assets that are closely linked within online user and topic networks are also likely to be highly correlated in their returns – this analysis offers insight into economic themes and investor preferences which drive asset co-movements.

The following chapter delves into the mechanisms shaping individual perspectives. I use two different instrumental variable approaches to document that user sentiments about assets' future performances are motivated, in part, by peers as well as by recent asset performance. I subsequently develop a theoretical case for how individual sentiment dynamics can drive changes in asset prices. I present several mathematical models, including a model with market-clearing, a dynamical systems model, and a model for bubble formation in the financial markets, which all demonstrate how 'social investing' can have market impact.

The thesis then uses a mixture of mathematical and econometric tools to empirically estimates the effect of WSB activity on asset prices in several ways. I document: higher retail trader demand following WSB discussions, the price impact of idiosyncratic investor sentiment in viral content, and the significant activity of WSB users around bubble-like stock price run-ups. The thesis concludes that the unique features of social

media, such as its scale and ability to facilitate the spread of information, make it an important market driver.

After making a strong case for the importance of human behaviour within the economy, I strive to better understand evolving real-world network dynamics – the next section of the thesis presents both a new dataset and novel temporal clustering technique targeted towards better understanding the dynamics of turbulent networks. The thesis analyzes the evolution of several social and political movements using data from online Reddit forums (r/BlackLivesMatter, r/Brexit, r/climate, r/democrats, r/Republican) and the development of international relations through studying international events data.

The key contributions of the thesis are analytical (developing several models to understand the macro impact of individual behaviour), empirical (extracting signal from noisy social-media data and leveraging network / causal frameworks to understand effects) and methodological (extending the network and data science toolkit to drive my analysis). The mathematical analysis within this thesis allows me to make a strong case for how social media activity has evolved over time and impacted society, as well as providing additional tools, datasets and algorithms to facilitate future research.

Acknowledgements

The PhD, as many parts of one's life, is an adventure: a time to explore the unknown and an opportunity to learn new things. However, what makes the journey worthwhile is the people one meets along the way – those who inspire us, as well as the friends who stay by our side through all of the challenges.

In the summer of 2018, I came across the research profile of a person who seemed like an explorer; someone with immense intellectual curiosity and energy – Doyne Farmer. As an aspiring researcher, I was thrilled to get a chance to have a video call and later in-person chat about various research ideas and interests. Little did I know that Doyne would become a mentor, advisor and supporter of my research – a key figure throughout my PhD. In early 2019, I was thrilled to be accepted to Oxford with the opportunity to do my DPhil under the supervision of Doyne. I could not have asked for a better supervisor – someone with great subject-matter expertise and diverse interests. Doyne is someone who inspires his students through his relentless ability to overcome many challenges: from starting a company, to publishing top research papers and writing a fantastic book (all at the same time). I am so grateful to him for this amazing opportunity.

Shortly after arriving at Oxford, I had the great fortune of meeting two other very important contributors to my research journey: my supervisors Xiaowen Dong and Renaud Lambiotte.

In addition to being a world-leading expert in deep learning on networks (making him a deeply knowledgeable and insightful supervisor), Xiaowen approaches supervision with dedication and great care for his students. Even from before I became Xiaowen's student, he was willing to spend time discussing research directions and helping me understand the latest literature. After I became his supervisee, Xiaowen prepared a semester-long course for just one other student and I to introduce us to the literature and techniques of Graph Neural Networks. Despite Xiaowen's students having diverse interests, he goes out of his way to tailor his supervision to them. For example, towards the penultimate year of my PhD, Xiaowen recommended that I travel to MIT and helped me seek out a fantastic research opportunity. Xiaowen's support and mentorship were critical parts of my DPhil experience – ones for which I am very thankful.

Upon arriving at Oxford, Renaud quickly became an inspiration for my PhD as someone with vast knowledge of network techniques, as well as a very kind and approachable member of the mathematics department. I felt inspired by Renaud to study new methods and explore. Renaud goes out of his way to say something positive and uplifting to his students, while steering them towards doing meaningful and insightful research. Simultaneously, despite being a sought-after academic, Renaud is the embodiment of intellectual curiosity – always humble and looking to learn something about a novel area. During my time on the academic job market, Renaud found time to offer guidance and encouragement, as well as forwarding me interesting positions and highlighting my potential strengths. I am very grateful to his supervision and meaningful contribution to my research journey.

I would be remiss not to mention another important contributor to my research journey

during the PhD – Sandy Pentland. In the spring of 2023, I traveled to MIT to broaden my research scope and learn from Sandy as part of the Human Dynamics group at the Media Lab. The experience at MIT was completely different to Oxford, and an invaluable contribution to the PhD adventure. I remember that every time I would enter the Media Lab, I would marvel at all of the different fields and new technologies being explored in one place. Sandy taught me how to think and dream broadly about the implications and scope of my research – all of our meetings, as well as the collaborative and interdisciplinary environment at the Media Lab, allowed me to consider the relevance of my work for policy, industry and innovation. Learning from Sandy helped develop my thinking as a researcher, fueled my curiosity and empowered me by showing me how many diverse interests and areas an individual can contribute to.

Finally, I would like to thank my examiners for taking the time to consider my thesis. Both Mirta and Pete are distinguished and pioneering researchers in their fields who have my greatest admiration; it is an honor to have my work examined by such esteemed experts.

Friends, Family, Colleagues As many an adventure book might tell us, the protagonist's quest is bound to fail without their trustworthy companions. If, therefore, I consider my PhD an adventure, it would have been a doomed endeavour without the friends and family that have supported me along the way.

I would like to thank my amazing partner in life and in research, Julian. Julian – you have inspired me and challenged me intellectually, encouraged me through the difficult times and been there to celebrate the successes, while always striving with me to be better. As my most trusted confidante and the kindest person I know, I would like to dedicate this work to you.

My family has contributed greatly both to my PhD and to the person that I am today – Mikhail, Elena, Valentina, Vsevolod and Eta. Parents, you mean so much to me: from helping me think through every-day challenges, to seeking out every opportunity to take us out for a wonderful celebration. You have been such a valuable support throughout my research journey and beyond; words cannot capture all of the things you have taught me. I believe that all good parts of me – the friend that people cherish and the researcher that colleagues value – originate in you.

While mentioning family, it would be remiss not to say a huge thank you to Julian's parents, Matthias and Yasemin, for their amazing contribution to my time in the UK and during the PhD. Matthias and Yasemin kindly welcomed me into their home, always preparing wonderful treats for me and Julian, and taking great care of us. I am extremely grateful.

The DPhil is a long journey – one that I cannot imagine doing without fantastic friends, mentors and colleagues at Oxford, as well as my lifetime friends from home. I would like to thank Mirta and Camelia – both fantastic researchers who I look up to and admire. Cheers to all of my INET and OMI colleagues who have welcomed me and shaped the way I think: from Maria, Blas, Toni, Marco, Alissa, François, Torsten, Jangho and Kieran who welcomed me upon my arrival to Oxford, to Sam, Aymeric, José, Ben, Valentina, Dragos, Bill, Pierre, Alex T, Bassel, Stefan, Cameron and Dorothy who made the work days so much more fun

and full of joy, as well as providing me with fantastic advice throughout the journey! A huge thanks to my neighbors both in Oxford (Jane and Martin) and on Boar's Hill (Emma, Robert, Thomas and Edward) for their kindness and taking me in as family. A shout out to my fantastic St. Catherine's College crew and other wonderful friends: Chloe, Christoph, Giedre, Fede, Felice, Andrey, Stan, Marnie, Fran, Jean-Guillaume and Juliette. Finally, a huge thanks to all of my friends from home: Ish, Bonnie, Michael, Paul, Kristín, Vinny, and Christina who came to visit and support me during the PhD and who have gone on some amazing adventures with me over the years, Ariel and Raghav who selflessly offered me their home during my time in Boston – and Ashley, Marina, Eva and Bernie who immediately made the city feel like home.

Contents

Statement of Originality and Completeness

This thesis contains work that was pursued in collaboration with my supervisors, other students and outside parties. I am extremely grateful for their input, and take the opportunity to outline their contributions and express my gratitude below.

Julian Winkler and I began to discuss and think about the impact that social media can have on the financial markets towards the beginning of my PhD in early 2020. The topic captured my attention as a phenomenon with potential importance, as retail traders turned to social media and investing in the wake of the COVID-19 pandemic. Julian was the perfect coauthor who complemented my knowledge of mathematical and data scientific methods with a sophisticated knowledge of the economic literature. The collaboration resulted in two papers which form the basis of chapters 4,5,6 (one currently submitted, the other in final pre-submission review). Julian and I kept constant communication, and it is challenging to pinpoint exactly who did what. However, much of the design for the empirical experiments pinpointing peer effects in their current form (Chapter 4), as well as the GIV and TAQ analysis (Chapter 6) are my contribution. The modeling sections evolved through joint thinking.

After presenting my research on social media (the `r/WallstreetBets` forum specifically) and its impact on the financial markets, the topic attracted a talented group of collaborators from the Oxford Man Institute (Bill Wildi, Dragos Gorduza, Xiaowen Dong and Stefan Zohren). Xiaowen and Stefan provided valuable guidance for the work. I use an excerpt of our jointly authored paper, where I act as a lead author (published in Journal of Portfolio Management), as a part of Chapter 3.

Finally, as part of my PhD, my goal was to study social movements and better understand the evolution of temporal networks and relationships. My interest was shared by a talented collaborator, John Pougué-Biyong, as well as my fantastic supervisors Renaud Lambiotte and Doyne Farmer. Our joined interest resulted in us writing two papers which underpin the research in Chapter 7: one currently published as part of NeurIPS proceedings and another in final stages of preparation for submission. In the first paper, John and I both designed the experiments and relevant statistics to extract from the data, Scale AI (a company specializing in AI data processing) provided data annotation and Renaud / Doyne offered excellent guidance. For the second paper, John and I both contributed to the algorithm, while I studied the parameters and formulated empirically-driven strategies that made the algorithm usable for the datasets presented in my thesis. Ultimately, John pursued a distinct set of experiments to mine in his thesis – the work presented in my thesis (which is centered on the application of the algorithm to for several specific settings) is my own. As a leading expert on clustering algorithms, Renaud helped shape the algorithm and guide our thinking on the project.

Out of the utmost respect and gratitude to these valuable contributions, the thesis proceeds in fist-person, plural tense.

Chapter 1: Introduction

Social forces and global economic and political trends are strongly interconnected. People's interactions and convictions influence their voting patterns and investment decisions. The political climate, in turn, impacts monetary policy and trade. Investment provides the economic fuel for companies to thrive and expand. This thesis uses a mixture of mathematical and econometric tools in order to better understand the granular dynamics of financial markets, political movements, social interactions, and the interplay between them.

Several factors make it an opportunistic time to contribute to this field using a mathematical, complex-systems toolkit. Social media acts as a communication and coordination platform – enabling individuals to come together at an unprecedented scale and with incredible speed. As more people became users of online platforms in the early two thousands, social media became a hotbed for political activism – later termed 'online activism' (Cammaerts 2015). More recently, these platforms have enabled individuals to come together and destabilize the economy. Consider, for example, the GameStop short squeeze, when coordination among millions of online discussants enabled them to unseat an investment giant.[1] The Silicon Valley Bank Run, on the other hand, demonstrated how social media could fuel downside panic.[2] The silver lining is that social media allows us to operate in a data-rich landscape, providing opportunities for novel research and quick regulatory action. Specifically, leveraging social media data, in combination with novel mathematical tools and methods, enables us to uphold and debunk long-standing theories of how people behave and influence each other, and to explore data-driven frameworks for the interplay between individual agents and the economy. In this thesis, I study a relatively unexplored

[1] https://www.nytimes.com/2022/05/18/business/melvin-capital-gamestop-short.html

[2] https://www.theguardian.com/business/2023/mar/16/the-first-twitter-fuelled-bank-run-how-social-media-compounded-svbs-collapse

dataset (activity from the infamous WallStreetBets subreddit forum) and offer a unique, mathematical and data-science perspective, which complements existing work (as discussed in my literature review).

Figure 1.1 describes the structure of the thesis. I begin by applying a data-scientific lens to the novel dataset tracking investor activity on social media in Chapter 3. The following chapters (4, 5, 6) rigorously analyze the connection between social dynamics online and financial stability. Chapter 7 broadens the scope of the thesis to study other online communities.

The contribution of this thesis is methodological, analytical and empirical. For the analytical component, I adopt a dynamical systems perspective and study the mathematical processes that explain how social dynamics online can drive financial fluctuations. I present three different approaches in Chapter 5 which allow me to estimate the stability / cyclicality of a system with social investing (using the complex system toolkit), as well as evaluate the long-term equilibrium effects (using quantitative finance models). Reconciling my analysis with data required several methodological contributions, which allow me to extract signal from noisy, unstructured information. Chapter 7, for example, describes and implements an extension of a signed clustering technique to the temporal setting. Finally, the thesis applies data science tools, such as large language models (LLMs) and network scientific techniques, as well as econometric tools (such as the granular and network instrumental variable approaches) to reconcile my analysis and observed dynamics.

A large portion of the thesis is dedicated to studying investors that actively share information and seek the advice of their peers online – a group I refer to as 'hype' and 'social' investors. Specifically, I focus on the WallStretBets (WSB) subreddit forum. The goal of this research is twofold: i) to better understand how individuals make up their minds about assets based on information that they are exposed to online and market conditions, and ii) to

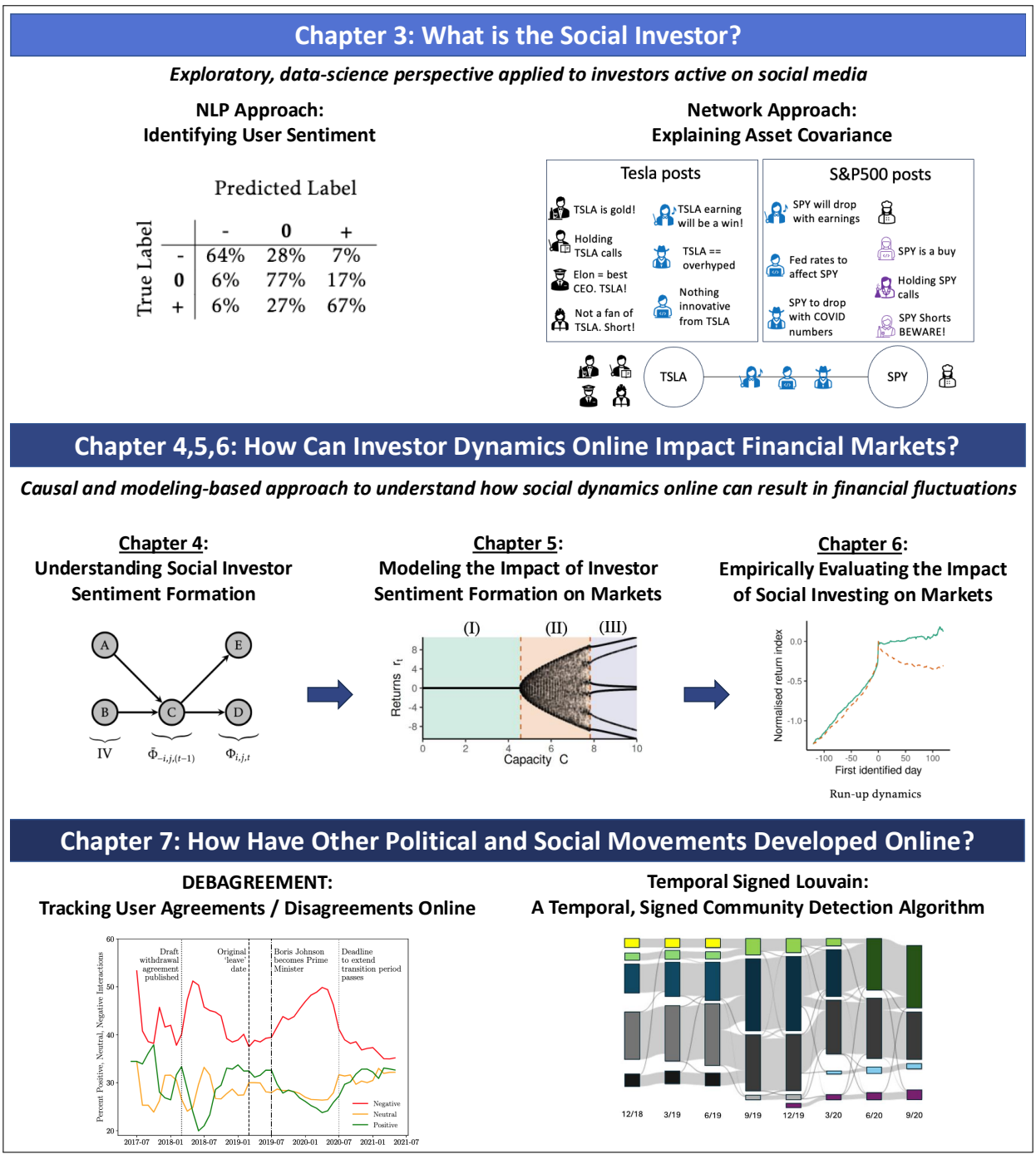


Figure 1.1: Thesis Road Map

study what properties of social media can lead to unique channels of market impact.

I begin with a text- and network-based analysis in Chapter 3. The thesis uncovers text-based and user-interest-based asset clusters within the WSB forum by applying LLM and network techniques to WSB submission and comment data. I subsequently explore the reaction of WSB users to market moves through an analysis of cumulative abnormal returns preceding and following submissions, and observe that users are generally reactive to price,

and that this pattern is particularly distinctive in 'meme' stocks.

Chapter 4 studies the way in which hype investors form their sentiments about assets (extracted using NLP methods) based on the information they are exposed to online and market conditions. The thesis leverages an Instrumental Variable (IV) approach to quantify the magnitude of peer effects among WSB users. I propose two separate identification strategies. The first is a temporal IV which leverages the historical sentiment of peers to predict their future sentiment about the same asset. The second is a network IV which uses the sentiments of 'friends of friends' in an interaction network. Across both specifications, I observe that peers play an important role in shaping individual investor perspectives. The results reveal that when the odds of peers expressing bullish over bearish sentiments double, the odds of a given user expressing bullish over bearish sentiment increase by an average of 15%. Overall, peer effects appear to play a greater role in sentiment formation than extrapolation (the tendency for individuals to extrapolate future returns from past performance).

An outstanding question is whether peer effects and social investing can, theoretically, have an impact on the financial markets. Chapter 5 presents several mathematical models for how sentiments and asset prices impact each other which incorporate my empirically-documented mechanisms behind sentiment formation (namely trend-following and peer effects). The models demonstrate that across many settings peer effects can play a substantial role. For example, in an equilibrium quantitative finance model with market clearing and liquidity provided by noise traders, I am able to show that heavy tails in attention paid to certain social media content can have market impact. In an adapted model of bubble formation (Barberis et al. 2018), peer effects manifest as a longer-memory process: I observe that a bubble takes a longer period of time to form and dissipate. Finally, in a complex systems model, I show that one can observe a range of behaviours in asset prices – from convergence to a steady state at zero, to periodic and quasi-period dynamics (bordering on

chaos) with multiple steady states.

If there are multiple mechanisms through which peer effects and online investor dynamics can impact asset prices, have we actually observed such changes in the financial markets? Chapter 6 leverages mathematical and econometric techniques to estimate the effect of WSB activity on asset prices in three ways. I document: higher retail trader demand following WSB discussions using Trade and Quotes (TAQ) data to identify retail trades (Boehmer et al. 2021), the price impact of idiosyncratic investor sentiment in viral content using a Granular Instrumental Variable (GIV) approach (Gabaix & Koijen 2023, 2021), and the significant activity of WSB users around bubble-like stock price run-ups (Greenwood et al. 2019). My findings suggests that it is likely that we have only seen the tip of the iceberg, in terms of the impact that social media can have on the financial markets and the economy more broadly.

After documenting the effect that hype investors have had on financial markets, the thesis broadens its scope and develops network theoretic / NLP tools to study the social dynamics in several other communities. I study the evolution of several political online forums (`r/BlackLivesMatter`, `r/Brexit`, `r/climate`, `r/democrats`, `r/Republican`) and the temporal dynamics of international relations, through the lens of international news events (Boschee et al. 2015). Due to the lack of high-quality, annotated data from social media the chapter begins by presenting a novel dataset tracking the evolution of agreements and disagreements in the online communities mentioned above. I use the data, annotated in partnership with Scale AI,[3] to better understand the properties and shortcomings of NLP algorithms on noisy social-media data. The thesis subsequently presents an signed, temporal clustering algorithm – an extension of modularity-maximizing methods on signed networks (Blondel et al. 2008, Traag & Bruggeman 2009). This algorithm is then applied to study the temporal clustering dynamics of the networks described above. My method allows me to gain insights into key differences between the evolution of discussions among Democrats

[3] `https://scale.com/`

and Republicans leading up to the 2021 US election, and to observe how country alliances have evolved over time.

The work presented in Chapters 3, 4, 5, 6 is based on several papers developed throughout the thesis, including Semenova et al. (2024) (published in Journal of Portfolio Management), Semenova & Winkler (2021) (under review at Management Science) and a manuscript in preparation for submission to Proceedings of the National Academy of Sciences. Chapter 7 is based on a dataset paper published at the Conference on Neural Information Processing Systems (Pougué-Biyong et al. 2021) and a paper published as a special prize issue for Rebuilding Macroeconomics (Pougué-Biyong & Semenova 2022). Out of the utmost respect and gratitude to the valuable contributions of my coauthors (detailed in the Statement of Originality) the thesis proceeds in fist-person, plural tense.

Chapter 2 contextualizes the thesis by providing a thorough literature review and highlighting the key contributions. Chapter 8 concludes and discusses future goals.

Chapter 2: Literature Review

This thesis uses diverse tools and cutting-edge techniques across several different areas. A large part of the analysis is underpinned by network theoretic and NLP techniques. In the network theoretic space, I build on and contribute to the network clustering literature, as well as the literature studying social networks in investor communities. In the NLP space, I review and contribute to the study and application of topic models and sentiment detection. My empirical methods leverage the latest developments in the peer effects in economics space, as well as identification methods for data exhibiting heavy tails. My modeling exercises extend quantitative models from dynamical systems and finance.

2.1 Networks

The simplest type of network (also known as a graph), $G = (V, E)$, is represented by a set of V nodes (vertices) connected by E edges (often represented as a node tuple) (Newman 2018). The structure of the network is often captured in the adjacency matrix A, which is NxN (where N is $|V|$) and contains a one in entry $A_{i,j}$ if nodes i, j are connected by an edge and zero otherwise. Advancements in network science have facilitated many more complicated representations of networks. Traditionally, networks are unsigned and undirected (resulting in a symmetric adjacency matrix). Directed networks capture the notion that a relationship may originate from one node and move to another, but may not be reciprocated, and can have a non-symmetric adjacency matrix: for example, trade relationships where a country may import goods from another but may not export to that country (De Benedictis & Tajoli 2011), or disease may spread from an infected patient to a healthy individual (Meyers et al. 2006). Edges may also capture different strengths of ties (where an edge may have a weight between zero and one) or different types of ties (Newman 2004). In the latter case, networks

can be *signed*, where edges may have a negative or a positive value associated with them – the negative edges may denote antagonistic / adversarial relationships, whereas positive edges may capture friendly ties (Leskovec et al. 2010*b*). Temporal networks, on the other hand, represent relationships that may be changing over time where edges have a temporal attribute (Masuda & Lambiotte 2016). More recently, the network literature has expanded to incorporate many different types of relationships and / or nodes in a multi-dimensional network or hetero-graph structure, where the graph $G = (V, E, D)$ may refer to a specific dimension D or the set of nodes V can be broken down into different node types (Kivelä et al. 2014).

In addition to exploring different types of relationships, a network may contain node features, which capture specific characteristics of the nodes – for example, if the nodes represent countries in a trade network, the node features may capture country-specific attributes, such as the Gross Domestic Product (GDP) or population. This development is of particular relevance for deep-learning on graphs, which has evolved into its own field focused on the study of Graph Neural Networks (GNNs). ML methods on graphs have largely focused on several tasks: node classification (assigning a label y_u to a node, $u \in V$, after observing the labels on a training set of nodes V_{train}), link / relation prediction (predicting whether an edge exists between two nodes after observe a set of existing and absent edges from E_{train}), community detection (given a graph with certain connections, determining the most similar clusters of nodes while maximizing 'dissimilarity' between communities and 'similarity' within communities) and graph classification (assigning a label y_G to graph G after observing the labels on a training set of graphs in G_{train}). The general architecture of many GNN methods takes in a graph with nodes, edges and node features; it subsequently learns a new node representation based on the initial node features of a node and its neighbors. Different approaches to learning the new node representation includes spectral methods (Hammond

et al. 2011, Kipf & Welling 2016, Defferrard et al. 2016) and spatial methods (Hamilton et al. 2017, Veličković et al. 2018, Monti et al. 2017). More recently, GNN methods have expanded to heterographs and multi-layer networks.

Within the network science literature this thesis focuses on several areas. Chapter 7 expands on existing network tools and presents a novel, signed, temporal clustering algorithm – for better context, I review the relevant literature and background in Chapter 7. Chapter 4 focuses on peer effects in social, economic networks – I review these in greater detail and outline the key contributions below. Finally, this thesis discusses how several future directions leverage deep-learning methods on networks.

2.2 Empirical Methods for the Study of Peer Effects Among Investors

The study of peer effects and the spread of information within social networks emerged as an important topic of research within the network science space after several studies documented that important life outcomes could be substantially impacted by the influence from an individual's peers. Christakis & Fowler (2013) review several of the authors' seminal works studying the spread of health outcomes (such as smoking, obesity and depression), social decisions (such as divorce) and tastes (such as music preferences).

A prominent strand of literature highlights the importance of peers and narratives in shaping investor perspectives. In his seminal work, Shiller (1984) discusses the excess volatility in stock prices relative to dividends. Since then, 'narrative economics' has played an increasingly important role in our understanding of investor decision-making and market moves (Shiller 2017, Hirshleifer 2020). Social media forums have been used to uncover various aspects of investor behaviour, such as: the myForexBook platform to study the disposition effect (Heimer 2016), StockTwits to study echo chambers (Cookson et al. 2022), and SeekingAlpha to study impression management and the propagation of noise (Chen &

Hwang 2022), to name a few (see Kuchler & Stroebel (2021), Cookson et al. (2024), Semenova & Winkler (2021) for a comprehensive overview). The WallStreetBets forum has been used to study: the interactions of different types of investors (Hu et al. 2021), the informational content of posts for prediction (Bradley et al. 2024), and the dynamics behind the GameStop short squeeze (Mancini et al. 2022). I study a different question to those listed above and propose an empirical methodology to more precisely identify how information shared by peers contributes to sentiment formation, as well as the impact of market surprise and reinforcement. Beyond studying a distinct question in the literature, I explore three different avenues for market impact.

A growing literature attempts to address concerns surrounding 'confounding variables' – variables which affect both the predictors and the dependent variables, altering the observed relationship between the two. An important strand of literature attempts to distinguish between peer effects (where a node in a network influences or causes outcomes in another), homophily (when similarities between nodes drive the same outcomes) and common shocks (shared influences may affect the outcomes for all nodes) (Aral et al. 2009, Angrist 2014). The impact of peers has been studied in several different financial contexts such as: the diffusion of micro-finance decisions in a social network (Banerjee et al. 2013), the effect of peers on risk taking (Lahno & Serra-Garcia 2015), the effect of social networks on saving (Breza & Chandrasekhar 2019), and the role of 'social learning' versus 'social utility' in financial decision-making (Bursztyn et al. 2014). By studying a broader set of investors in a natural experiment, my research question is similar to Pool et al. (2015), who demonstrate that socially connected fund managers hold similar stocks. Social media data, however, presents a novel identification challenge. Fortunately, other papers in economics offer promising methods to leverage naturally occurring variation in peers for identification. An area which pioneered many of these techniques investigates peer effects in the classroom (see Epple &

Romano (2011), Sacerdote (2011) for a general overview, and Duflo et al. (2011) for a prominent example). Social networks are also an active area of study (see Bramoullé et al. (2020) for a recent review). This thesis highlights how to transfer well-established techniques from the empirical peer effects in the classroom and networks literatures to social media data.

2.3 Asset Returns & Social Media

Empirical Investigation Several papers have documented that investor discussion forums can have predictive power for market returns (Antweiler & Frank 2004, Sabherwal et al. 2011, Chen, De, Hu & Hwang 2014, Azar & Lo 2016, Agrawal et al. 2018). This thesis quantifies additional channels through which social media has had an impact. I leverage the methodology of Boehmer et al. (2021) to show how retail trading volumes follow the attention paid to specific assets on the forum, complementing other works tracking WSB positions, such as Welch (2022). I present a new channel for price impact and use a GIV approach to demonstrate how virality online may be driving volatility in the financial markets (Gabaix & Koijen 2023). Since its development, the GIV methodology has been used to study several important questions, including credit risk (Galaasen et al. 2020) and capital flows (Gabaix & Koijen 2021). This is the first work, to my knowledge, to study the impact of social media and viral content. The GIV approach allows me to study fluctuations in returns at the weekly time-scale, however, I also consider that social movements can have more long-term consequences, namely through driving asset price bubbles. Several papers empirically study the formation bubbles in the markets. Greenwood et al. (2019) propose a methodology to study price run-ups in sectors. I adapt their framework to my setting and demonstrate the increased activity of hype investors in assets that exhibit bubble-like dynamics. My exploration prompts important questions for further research. Recent studies use investor-level holdings data to track positions and profits / losses of individuals (Pearson

et al. 2021, Balasubramaniam et al. 2023): given the quickly growing hype investor population in the market (a statement justified in the thesis), there is a pressing need to study these individual investors' profit and loss profiles as the re-distributional consequences of social influence among investors (An et al. 2022).

Models In the presence of social dynamics, what behaviour would we expect for asset prices in the financial markets? This thesis presents three models and discusses the key insights: i) a model where strategic complementarities among investors motivate them to share information online, ii) a model for asset price bubbles with extrapolative and social investors, and iii) a complex systems model with different stable and unstable emergent regimes. I review the relevant literature underpinning my analysis below.

The economic interest in asset mispricing has a long history, with examples dating back to Tulipmania in the Netherlands in the 17^{th} century (Garber 1989). Of particular relevance to this thesis are frameworks analyzing: the spread of information (Veldkamp 2006), strategic complementarities (Hellwig & Veldkamp 2009, Zenou 2016), and psychological models, such as diagnostic expectations (Bordalo et al. 2021) / extrapolation (Glaeser & Nathanson 2017). Models for the impact on observable network ties on asset price fluctuation (for a prominent example see Golub & Jackson (2010)) are less relevant for this thesis, as we focus on the *average informational content* shared by peers. I, therefore, adopt a model framework with strategic information complementarities and extrapolation. I also demonstrate how my estimates can be used in conjunction with existing models for bubbles to understand the role that social dynamics play in bubble formation (Barberis et al. 2018, Hirshleifer 2020). The goal of extending existing models to the current setting is to: i) explain a setting in which investors share their strategies, ii) justify the observed negative relationship between investor sentiment and returns, iii) explain channels through which investor information sharing online can impact markets – such as through heavy tails in the popularity of certain

content, or through the formation of asset bubbles.

A different strand of literature in the computational social sciences aims to model online collective dynamics (Bacaksizlar Turbic & Galesic 2023, Jusup et al. 2022, Nguyen et al. 2012, Becker et al. 2017, Kubin & Von Sikorski 2021), with several works focusing on investor forums specifically (Mancini et al. 2022, Lucchini et al. 2022). Researchers have also scrutinised the interactions of heterogeneous agents in stock markets, following many, sometimes adaptable, rules(Lux 1998, Hommes 2021), with some papers focusing on social dynamics specifically (Kirman 1993, Avery & Zemsky 1998, Cont & Bouchaud 2000). A highly relevant branch of literature considers market evolution and market ecology as a way to understand changing market conditions (Scholl et al. 2021, Farmer 2002). However, no studies have attempted to apply a complex systems approach to better understand the emergent financial dynamics driven by online investor discussions. This thesis presents an analytic approach and empirical methodology to help address this gap.

2.4 Text as Data

Textual analysis has long been viewed as an important analytical method and data extraction technique. In this thesis, I focus on three important areas of development: sentiment and stance detection, topic modeling and economic applications. I contribute to the field of stance detection in Chapter 7 where I analyze the performance of various NLP methodologies when applied to stance detection to noisy social media data. I observe that state-of-the-art deep-learning architectures fail to perform well at this task and develop a dataset designed to train and improve these models. My analysis of social media and finance tests several tools for sentiment detection in the financial, social media contexts and leverages the output as an important component underpinning the rest of my analysis. A substantial part of this thesis was developed before the prevalence of ChatGPT (OpenAI 2023) and later

LLMs (Jiang et al. 2023), however, I run several tests on this latest group of models and also discuss their performance.

Topic Models Topic models are often employed to map out the distribution of information in a textual corpus (Blei et al. 2003, Angelov 2020). In economics and finance, they are used to uncover the latent discursive directions that individuals can follow when expressing an opinion, which allows researchers to quickly detect salient points in the discourse and use them as signals for downstream analysis (Gentzkow et al. 2019). However, prior work in the financial / social media space leverages older topic models which do not make full use of the new modelling capacities offered by large language models (LLMs) (Schou et al. 2022). In this thesis, I demonstrate how latest developments in the LLM space, in combination with network tools, can yield novel insights.

Sentiment Detection A similar, important task that has gained attention is extracting individual sentiment from a piece of text. The goal of this thesis is specifically focused on classifying finance-related social media posts with respect to the author's outlook on an asset: bullish, bearish or neutral. Previous papers have either relied on counting keywords (for example, 'buy', 'sell', 'call', 'put' etc. Buz & de Melo (2021), Bradley et al. (2024)), or they have used off-the-shelf rule based sentiment classifiers (Wang & Luo 2021), such as VADER (Hutto & Gilbert 2014). The advantage of such approaches is their simplicity, but they have several problems. Choosing keywords/phrases is inherently arbitrary – such an approach could easily miss posts that are clearly positive/negative because they don't contain the required keywords, and it is vulnerable to mistakes when authors negate phrases in the dictionary (e.g. 'calls are going to be worthless'). Algorithms like VADER, by contrast, lack the specific domain vocabulary of the WSB forum, hence their performance can be quite poor (Wang & Luo 2021). To resolve these issues, I leverage a deep-learning, fine-

tuning approach from the transformers-based pre-trained BERT model (Devlin et al. 2018, Araci 2019).

Chapter 3: The Social Investor: What is WallStreetBets?

3.1 WallStreetBets Overview

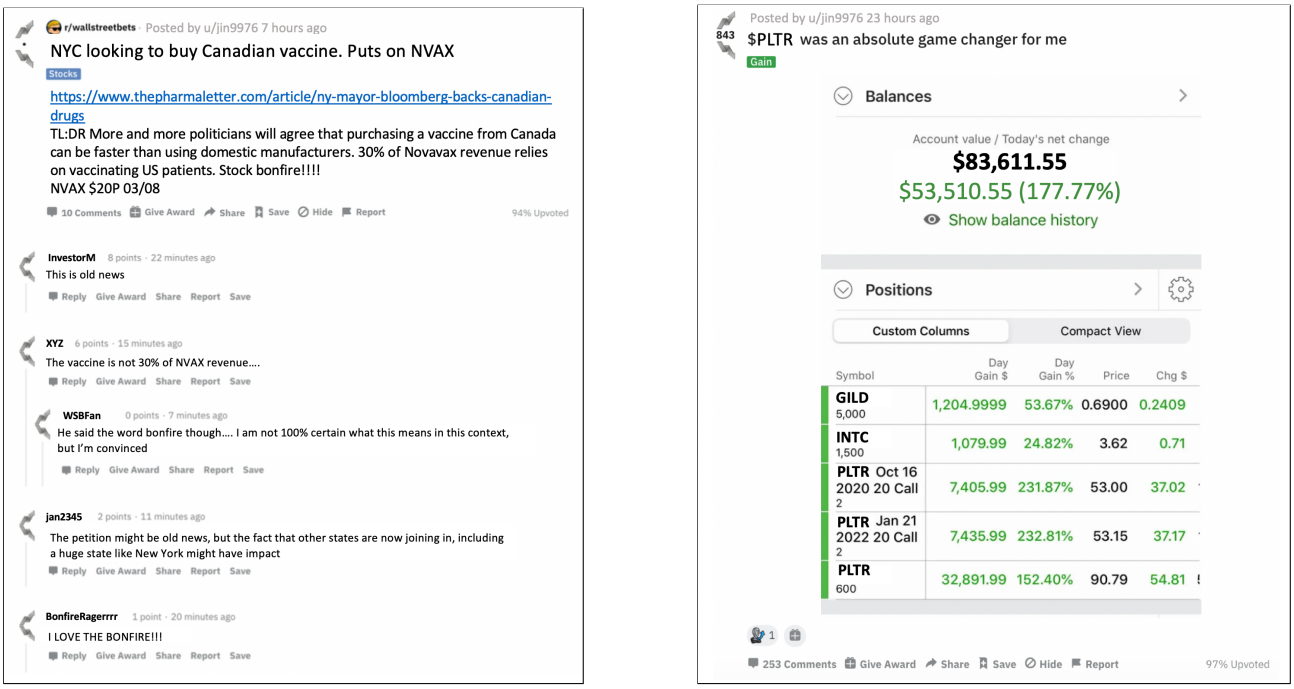
Reddit, launched in 2005, is a social news aggregation, web content rating, and discussion website. It was ranked as the 8th most visited site globally in March 2024,[1] with over 268 million anonymous, active, weekly users.[2] The website's contents are self-organised by subject into smaller sub-forums, 'subreddits', which discuss a unique, central topic.

Structure of WSB Within subreddits, users publish titled posts (called 'submissions'), typically accompanied with a body of text or a link to an external website. These submissions can be commented and 'upvoted' or 'downvoted' by other users. A ranking algorithm raises the visibility of a submission with the amount of upvotes it receives, but lowers it with age. Therefore, the first submissions that visitors see are i) highly upvoted, and ii) recent, with the precise algorithm considered private intellectual property.[3] Comments on a submission, visible to anyone, are subject to a similar scoring system, and can, themselves, be commented on.

Features The WSB subreddit was created on January 31, 2012, and reached one million followers in March 2020.[4] As per a Google survey from 2016 (later reinforced by more recent surveys[5]), the majority of WSB users are 'young, male, students that are inexperienced investors utilizing real money (not paper trading); most users have four figures in their trading account'.[6] Individuals on the forum discuss and express their sentiments about

[1] https://www.similarweb.com/top-websites/
[2] https://backlinko.com/reddit-users
[3] https://www.reddit.com/r/help/comments/7l7686/order_of_posts/
[4] https://subredditstats.com/r/wallstreetbets
[5] https://www.reddit.com/r/wallstreetbets/comments/a2qvj3/wsb_demographics_survey_results_a_portrait_of/
[6] https://andriymulyar.com/blog/how-a-subreddit-made-millions-from-covid19

stock-related news. In addition to market discussions, there is ample evidence of users pursuing the investment strategies encouraged in WSB conversations. Users post screenshots of their investment gains and losses, which subreddit moderators are encouraged to verify – a dynamic reminiscent of Shiller's (2005) description of an asset bubble. The discussions are whimsical, but mostly investment-focused.



(a) A typical discussion on WSB

(b) A sample screenshot of user profits

Figure 3.1: **What does WSB look like?** These snapshots display typical discussions on WSB. The exact text, usernames, and conversation details have been modified to protect user identities.

Figure 3.1a displays a typical exchange on the WSB forum: individuals discuss stock-related news and their sentiments on whether this will affect stock prices in the future. In addition to market discussions, there is ample evidence of users pursuing the investment strategies encouraged in WSB conversations. Users post screenshots of their investment gains and losses, which subreddit moderators are encouraged to verify, as illustrated in Figure 3.1b.

Available data We downloaded WSB data using the PushShift API.[7] PushShift records all comment and submission data at the time of creation. We collect data from the inception of the forum in April 2012 up to 24 June 2022. The full dataset consists of two parts. The first is a total of 1.4 million submissions, with their authors, titles, text and timestamps. The second is comprised of 16.5 million comments, with their authors, text, timestamp, and the identifier of the parent comment or submission. Submission and comment numbers have grown exponentially since 2015.

For my analysis in Chapters 4, 5, 6, I rely on shorter sample of data spanning from January, 2012 to July, 2020 – consistent with the time period when the analysis was performed (and later reinforced by experiments on the full dataset spanning mid-2022). Importantly, the results of this analysis do not hinge on the 2021 GameStop (GME) short squeeze.

A key remaining question is the broader relevance of this data and continued behaviours similarly to the ones exhibited on WSB. We present evidence of the increased popularity and continued importance of the social trader in Appendix 3.A.

Identifying assets The following chapters predominantly rely on submissions for text data, since they are substantially richer than individual comments. Comments are used to trace interactions between discussants. In order to understand how users discuss specific assets, we extract mentions of *tickers* from the WSB submissions' text data. A ticker is a short combination of letters, used to identify an asset on trading platforms. For example, 'AAPL' refers to shares in Apple, Inc.

Sentiment model Our goal, with regards to the text data in WSB, is to gauge whether discussions on certain assets express an expectation for their future price to rise, the 'bullish' case, to fall, the 'bearish' case, or to remain unpredictable, the 'neutral' case. Among

[7] https://pushshift.io/

		Predicted Label		
		-	**0**	**+**
True Label	**-**	64%	28%	7%
	0	6%	77%	17%
	+	6%	27%	67%

Table 3.1: **Fine-tuned FinBERT confusion matrix:** We use 10% of our hand-labeled data to test the performance of FinBERT on out-of-sample sentiment prediction. The results highlight the model's ability to predict sentiment with reasonably high accuracy.

other alternatives, we pursued a supervised-learning approach to identify the sentiment expressed about an asset within a WSB submission. This required a training dataset, for which we manually labelled 4,932 random submissions with unique ticker mentions as either 'bullish', 'bearish' or 'neutral' with respect to the authors' expressed expectations for the future price. We used the FinBERT algorithm for labeling (Araci 2019) - a financially oriented modification of Google's Bidirectional Encoder Representations from Transformers (BERT) algorithm (Devlin et al. 2018).

We trained FinBERT on 75% of the labelled data, and used the remaining 25% for validation and the test set. Table 3.1 plots the out-of-sample confusion matrix. For the out-of-sample test, we train FinBERT on 75% of the available data and use 15% for validation; we then compute what the algorithm predicts for the remaining 10% of data. We achieve 70% accuracy on the test set. This is better than a LASSO regression's accuracy or sentiment dictionary, which were implemented separately and are not cover here. We test several other more recent large language models, for example Mistral 7B (Jiang et al. 2023), and do not observe in a significant improvement in performance (though context-enhanced LLMs, such as Zhu et al. (2021), have potential to improve the performance - we leave this for further research). Here we present the results of our data analysis preceding the GME short squeeze; The analysis in Semenova et al. (2024) shows the sentiment extraction produces consistent results for the entire dataset spanning mid-2022.

Key sentiment variable The sentiment classifier assigns three probability scores to each submission about a ticker: the probability of a submission being bullish, $P(\phi = +1)$, bearish, $P(\phi = -1)$, neutral, $P(\phi = 0)$. The probabilities sum to one. At the time t when an author i posts about asset j, we use the probability scores above to calculate a continuous sentiment score between $(-\infty, \infty)$:

$$\Phi_{i,j,t} = \frac{1}{2} \log\left(\frac{\mathrm{P}(\phi_{i,j,t} = +1)}{\mathrm{P}(\phi_{i,j,t} = -1)}\right). \tag{3.1}$$

Submissions labeled as bullish ($\mathrm{P}(\phi = +1) = 1$) (or bearish ($\mathrm{P}(\phi = -1) = 1$)) are set to $\mathrm{P}(\phi = +1) = 0.98$ (or $\mathrm{P}(\phi = -1) = 0.98$) to retrieve a finite value for the log-odds. For certain exercises, we assign three categorical variables (bullish, bearish, neutral) to a submission's sentiment, encoded with a one if the corresponding label received the highest probability from our classifier. As such, this categorical variable $\phi^{+1}_{i,j,t}$ will be equal to one is author i's post about asset j at time t is categorised as bullish; $\phi^{0}_{i,j,t}$ and $\phi^{-1}_{i,j,t}$ will be zero. We leverage these variables to investigate investor sentiment throughout the thesis.

We argue that the error rate and distribution of errors from our sentiment classifier are appropriate for the task we study. In most of our analysis, we use a continuous transformation of the sentiment – the log-odds of bullish over bearish sentiment. In our confusion matrix, we observe that the model mistaking bullish for bearish sentiment and vice-versa is rare: 7% in the former case and 6% in the latter case. The model has a greater difficulty distinguishing positive/neutral and negative/neutral, however, this is expected, given the continuous nature of language – fortunately, these errors have a lesser impact on our analytical result. We observe human annotator-to-annotator disagreement of 10%.

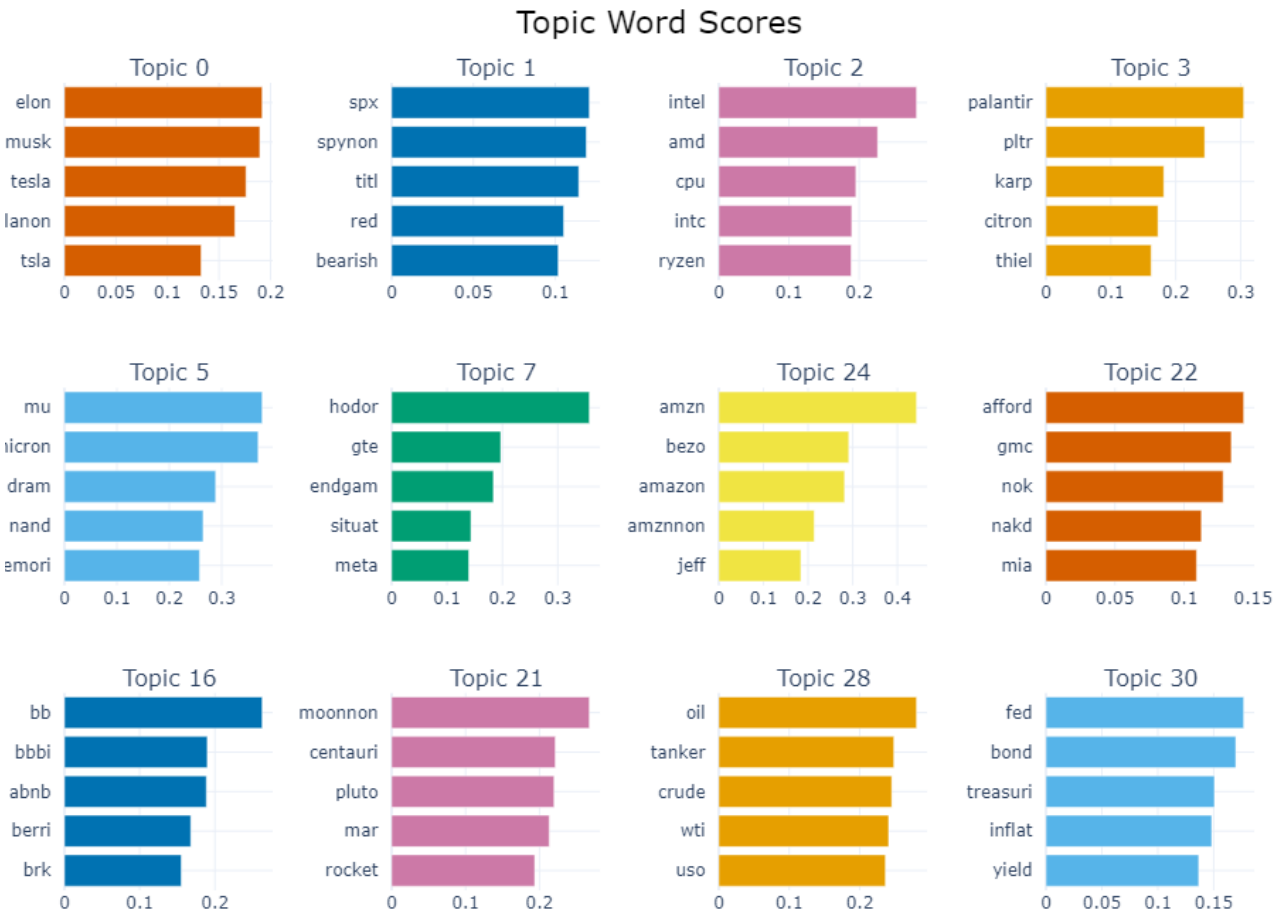
3.2 WallStreetBets Forum Dynamics

In this section, we characterise the discussion landscape of the WSB forum. We consider how assets are related on WSB in terms of topic and user interest. Additionally, we explore the relationship between WSB and markets through studying the returns preceding and following posts. As explained earlier in this section, Chapters 4, 5, 6 primarily rely on data spanning July 2020; this section consider the full dataset through June 2022.

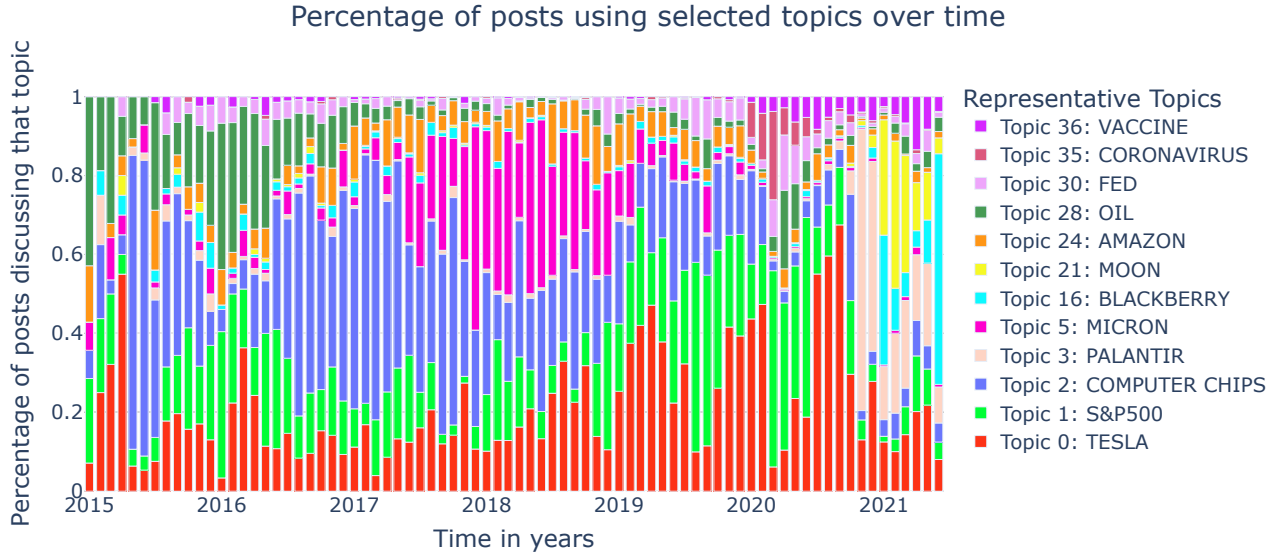
3.2.1 The topic landscape of WSB

We use the Bert LLM Topic package to perform topic modelling on WSB, which accounts for non-linear semantic similarities between texts which go beyond word-level co-occurrences (a limitation of traditional topic models) (Angelov 2020). This model gives us a mapping from each post in the sample to a learnt set of topics. Each topic has a unique set of representative terms associated with it, and posts are mapped to topics based on how well the text within the post matches that of a particular topic. Figure 3.2 highlights our key findings from employing the BERT topic model: Figure 3.2a present some of the key topics of discussion on WSB, while Figure 3.2b displays their prevalence on the forum over time. Each post within WSB is given a probability vector whose length is equal to the number of extracted topics, identifying the extent to which the post is associated with different topics. For example, a post which discusses pharmaceutical companies and how they will profit from the COVID vaccine will be given a higher probability of being associated with topics 35 (COVID) and 36 (vaccine approvals).

Figure 3.2b is constructed by first splitting our dataset into months and then calculating the fraction of the discourse dedicated to that topic within that month for the ten most popular topics. The span of topics covered here is a subset of the overall topics identified by the model, which serve to illustrate the key areas of attention for the forum across time. We



(a) Sample of Representative Topics Discussed on WSB; the figure displays topics and the frequency with which the most popular words appear within the topics (ranked by the categorical term frequency - inverse document frequency score, described in (Angelov 2020)).



(b) Topic Distribution over Time

Figure 3.2: **The WallStreetBets Discussion**; we present several representative topics from WSB in Figure 3.2a, as well as their relative importance within the forum in Figure 3.2b.

note that several topics and their prevalence follow news, such as CORONAVIRUS (topic 35) which appears in March 2021 and VACCINE (topic 36), which are more prevalent during the introductions of the first COVID-19 vaccines in end of the year 2021. The topic model also tracks the emerging focus of the forum on so called 'meme stocks' such as TESLA (topic 0), PALANTIR (topic 3), and BLACKBERRY (topic 16). Contrary to other transient topics, we detect a persistent interest of the forum in SPY (S&P500) as noted by the prevalence of topic 1, which represents attention toward the overall state of the market.

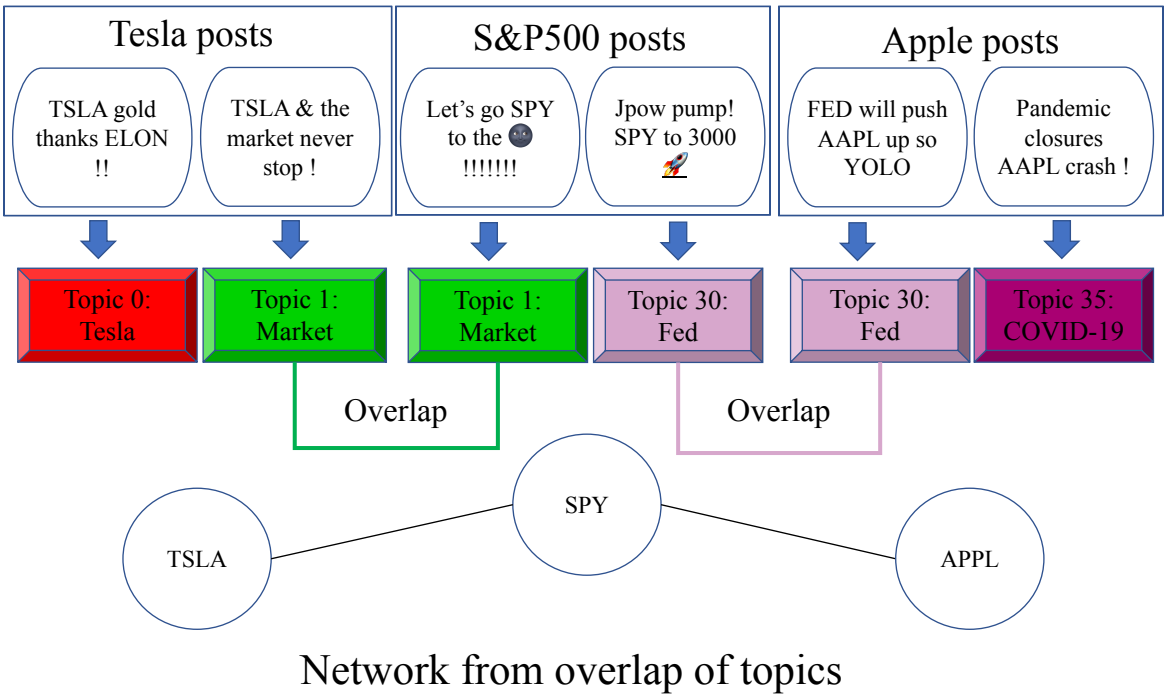
3.2.2 How are assets related to each other on WSB?

The WSB forum is uniquely suited to study the relationships between assets, as they are perceived and discussed together by retail investors. We choose two methods to map the relationships between assets and to create a ticker-to-ticker network structure: the *Topic Network* approach and the *Submission Network* approach. The *Topic Network* uses the frequency with which assets are discussed within the same topics to create inter-asset connections, indicating that investors perceive the assets to be influenced by the same general financial trends. The *Submission Network* tracks which groups of investors are interested in, and create submissions about, the same assets. We describe both approaches below.

Topic Network In the *Topic Network* approach, we link assets based on the frequency with which they are mentioned in the same set of topics. The intuition behind the approach rests on the idea that assets that are discussed within the same financial topics are likely related to each other through some underlying fundamental relationship.

We explain our intuition through a simplified example. Let us consider two assets: i) interest rate swaps and ii) bonds. We extract a topic from our topic model about the FED, and we observe that both bonds and interest rate swaps are frequently brought up in posts that are labeled with the 'FED' topic. This is unsurprising, given the fact that FED decisions

(a) *Topic Network* construction illustration



Network from overlap of topics

(b) *Submission Network* construction illustration

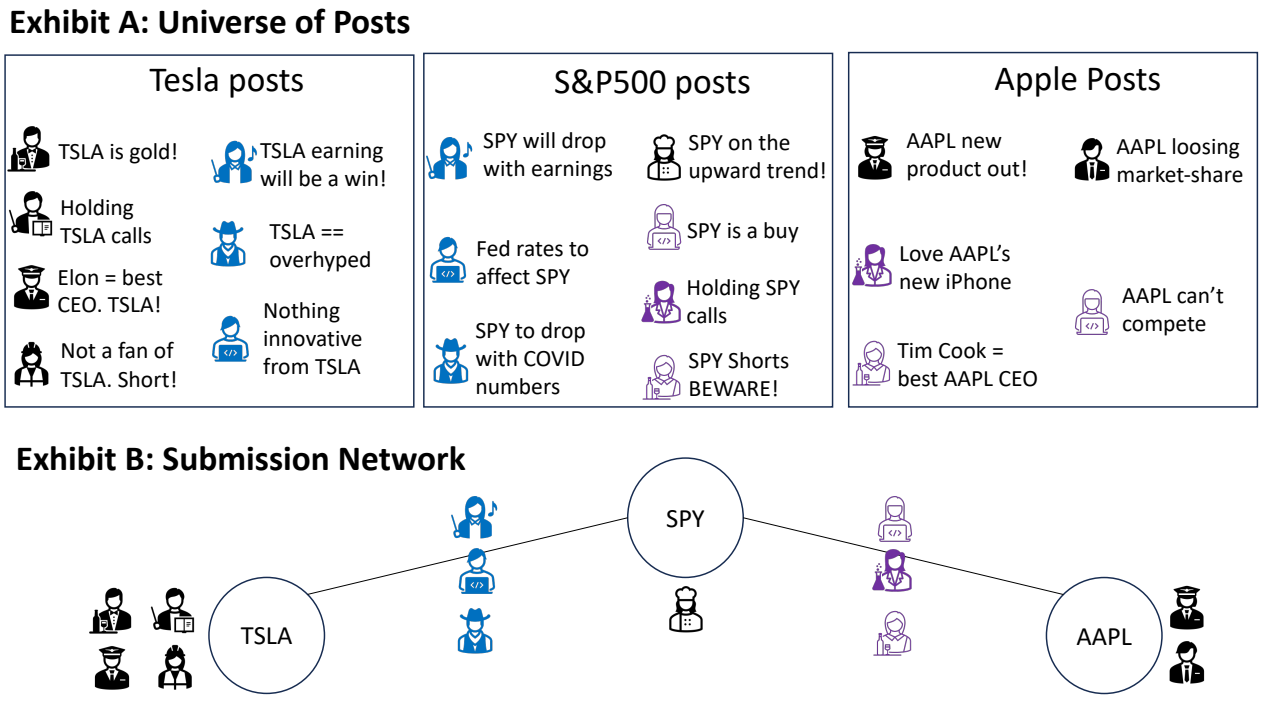


Figure 3.3: Network Construction; we demonstrate how links are identified between assets in the *Topic Network* and *Submission Network*.

would strongly affect the valuations of both bonds and interest rates swaps. In our *Topic Network* exercise we would place a connection between bonds and interest rate swaps, since

they are linked by the 'FED' topic. This intuition can be extended to better understand the meaning behind the tickers linked through out topic model: they are tickers that are frequently brought up under the same topics, indicating that they are linked through some underlying economic discussion theme.

Figure 3.3a provides a toy example. TSLA and SPY are connected because they are both mentioned in the topic discussing the overall stability of the market (*market* topic); SPY and APPL are connected since they are both discussed in conjunction with the *FED* topic.

In practice, we also add a thresholding step when constructing links between tickers based on overlapping topics. We select the top 50 topics, ranked by frequency of appearance in posts. This means that we only count a topic if it is represented often enough across our dataset. We then count all the topics that correspond to a given ticker and we keep the topics that are present in 20% or more posts related to that ticker – this is done to filter out less important conversations about assets. The filtering leads us to drop tickers whose topic distribution is very diffuse, meaning that discussions about it do not have consistency. As a prime example, this leads us to drop GME from our topic network as the topic distribution in GME posts is too dispersed.

Submission Network The community structure within WSB offers a different perspective on how assets are related within the forum. To construct our submission network, we look at the overlap of users who create posts about two assets. Two assets are linked if there is a sufficiently large fraction of users discussing both assets simultaneously.

Figure 3.3b provides a simplified example. We observe a subset of users creating posts about TSLA and SPY simultaneously, and a different group of users writing about AAPL and SPY. However, we do not observe the same users creating posts about TSLA and AAPL. Therefore, we create two links: TSLA / SPY and AAPL / SPY.

In practice, our network construction exercise is slightly more complicated. We create

weighted links between assets, normalized by the total number of users mentioning each asset. This means that if asset A is discussed by ten users, and all of them also post about asset B - we would create a link with weight one from asset A to B. However, if 10,000 users posts about asset B, we would not create a link back from B to A, since only a negligible fraction of posters about B care about A. Therefore, the *total weight* of the link between A and B would be 1 from the first created link. In practice, our threshold for including a link in our weight calculation is 20% – at least 20% of users posting about one ticker must be posting about the other ticker. Relating this back to Figure 3.3b, we observe that three out of seven SPY posters also post about AAPL (weight 3/7) and that three out of five AAPL posters post about SPY (weight 3/5) – the total weight for the link between AAPL and SPY would, therefore, be 3/5+3/7. We filter our submission network to contain tickers that are mentioned at least 150 times on the forum.

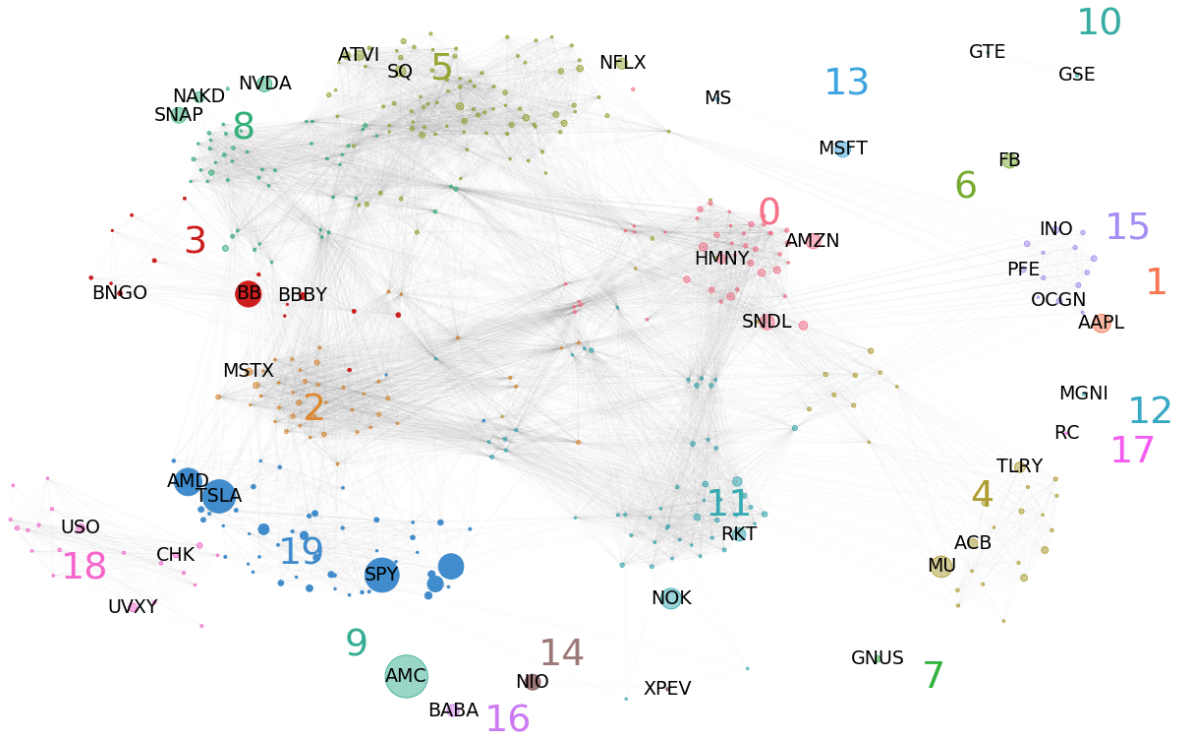
3.2.3 Asset network structures

Table 3.2: **Subset of Clusters extracted from the *Topic Network***

Cluster	Popular Tickers within Cluster	Posts	Total No. of Tickers
0	AMZN, SNDL, HMNY, MVIS, WISH, CLNE, CCL, DKNG, SUNE	12,373	41
1	AAPL	2,530	1
2	MSTX, AVXL, MNKD, OPK, AMDA, AUPH, SENS, AMRN, ADMP	3,312	44
3	BB, BBBY, BNGO, BRK, BBY, DB, BLNK, ABNB, BBW, BTT	9,898	12
4	MU, TLRY, ACB, CGC, APHA, CRON, AMAT, MO, OGI, HEXO	6,557	17
5	NFLX, SQ, ATVI, PLUG, WMT, SBUX, LULU, CRM, CMG, MCD	10,045	72
6	FB	1,580	1
8	SNAP, NVDA, NAKD, NAK, FSLY, TTD, DBX, KR, TWLO, GNC	5,763	35
9	AMC	19,686	1
11	NOK, RKT, WKHS, ZOM, QS, SRNE, TRCH, ASO, LFIN	11,349	33
13	MSFT, MS	1,849	2
14	NIO, XPEV	2,160	2
15	INO, OCGN, PFE, NVAX, MRNA, CVS, JNJ, AZN, VXRT, TEVA	2,474	13
16	BABA	1,175	1
18	USO, UVXY, CHK, WTI, XOM, SVXY, NAT, RIG, UCO, BP	1,001	12
19	SPY, TSLA, AMD, PLTR, SPCE, NKLA, QQQ, PRPL, GM, ROKU	28,438	30

Topic Network – Results Figure 3.4a presents our *Topic Network* – links indicate which two tickers are likely to be mentioned together within the same economic discussion topics. A list of tickers and their associated clusters are presented in Table 3.2. We observe that

(a) Topic network and extracted clusters



(b) Submission network and extracted clusters

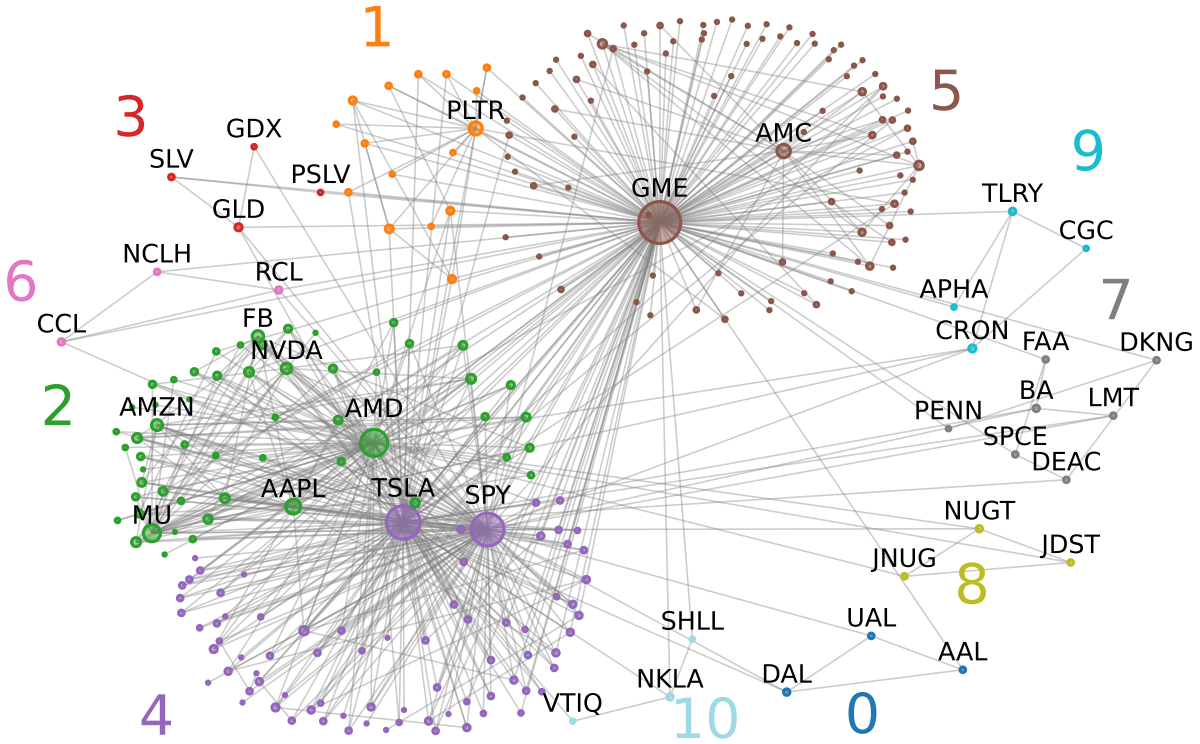


Figure 3.4: **Extracted Networks from WSB**; we illustrate the ticker networks extracted using both the *Topic Network* and *Submission Network* approaches. Clusters of closely connected tickers are extracted using the Leiden algorithm from (Traag, Waltman & Van Eck 2019) and highlighted with the same color in both networks. In Figure 3.4a, ticker nodes are scaled by the relative size of their number of posts; only the top 50 most mentioned topics are displayed. In Figure 3.4b, nodes are scaled by the number of connections they have.

larger and frequently discussed companies, such as AMC, have more cohesive topics which are not referenced in many other companies' posts. Therefore, the posts of several tickers are isolated from the main connected component of the topic network. These includes: AAPL (cluster 1), FB (cluster 6) AMC (cluster 9), and BABA (cluster 16). We interpret the isolation of these network components as the result of a systematic use of a *limited* and *distinct* set of topics when discussing these stocks. A post about AAPL will mobilise only topics similar to other AAPL discussions: focusing the attention of readers more narrowly on discussion points unique to Apple Inc as opposed to the wider market. Contrary to that, smaller companies tend to have a more central position in this network, and are often connected to stocks with higher market caps. This is visible by the more spread out cluster 19, which encompasses a variety of both smaller and larger companies as well as ETFs, likely connected through broader discussions about the economic climate.

Additionally, we find smaller clusters of assets related to a narrow set of economic themes within the market. A prime example of this is cluster 15, including INO, OCGN, PFE. These firms are likely linked by the same topics of drug discovery, FDA approvals and the COVID vaccine, in a way that is distinct to this particular group of assets.

We use our *Topic Network* as a distance mapping between firms that takes into account their 'discursive environments'. We interpret tickers belonging to the same cluster as an indication of greater similarity within their discussion topics. Consequently, belonging to the same cluster could be an indicator of a more correlated information set driving returns for all assets within the cluster.

Submission Network – Results Figure 3.4b presents the results of a clustering exercise on the *Submission Network*. We observe that similar assets appear to be mentioned within the detected clusters – implying that individuals self-select into discussions about certain asset categories. Said differently, a distinct group of investors is interested in marijuana

stocks to those discussing the cruise line industry. This is perhaps most pronounced in the small cluster of gold ETFs containing JNUG, NUGT, JDST – cluster 8. Several other smaller clusters, dominated by a niche sector, are also visible in clusters 0, 3, 8, 9. We observe a large cluster around the SPY ETF and TSLA (cluster 4), as well as a large 'meme stock' clusters around GME (cluster 5). We observe a distinct cluster with some large, popular tech stocks including AAPL, FB, MU, AMZN, AMD (cluster 2). In Appendix 3.B, we explore the correlation among assets in different clusters, as well as returns to WSB advice within the different clusters.

Returns and Correlations across Clusters In our *Submission Network*, we observe that distinct groups of investors are interested in different types of tickers: some users come to WSB in order to discuss pharmaceutical companies, while others are interested in trading natural resource ETFs, or airline stocks. For this reason, tickers clustered through the *Submission Network* tend to have returns that are highly correlated – clustered assets appear to have an average return correlation of greater than 0.2, while assets across our entire dataset appear to Figure an average correlation of 0.16. The most highly correlated asset returns appear to be within clusters 0, 6, and 9: the returns of the assets within these clusters Figure correlations of 0.80, 0.84 and 0.60, respectively. The average next-day returns of investing according to the extracted sentiments within submissions is statistically significant and negative across most clusters. The most negative average returns are exhibited within clusters 0 and 5; average returns of submissions in cluster 7, on the other hand, are statistically significant and positive. This demonstrates that most investors on WSB lack insights into future market moves. Notably, investors into the 'hype' cluster tend to lose the most. However, niche groups of investors may have worthwhile market insights – as demonstrated by returns for cluster 7. Appendix 3.B present returns to trading according to WSB advice and correlations across submission clusters.

A similar pattern can be observed from the *Topic Network* clusters. Investing according to the sentiments of submissions within most topic clusters results in a statistically significant, negative average next day return. However, similarly to our observations from the *Submission Network*, certain topic clusters Figure positive daily returns, on average: cluster 1 containing AAPL, 13 containing MSFT and MS, 14 containing NIO and XPEV (two electric car makers), and 16 containing BABA. We notice that smaller clusters with fewer assets perform better than much larger clusters, again demonstrating that the forum as a whole may lack market insights, however specific topics and groups of investors may have promising insights. In Appendix 3.B we present returns to trading according to WSB advice across topic clusters.

3.2.4 Returns preceding and following posts

What are the characteristics of assets returns before and after they are mentioned on WSB? Is there any evidence that WSB users can predict returns, or are they trend followers just like other retail investors?

Framework We consider how asset prices are changing shortly before and after posts on the forum. We perform two types of analyses: i) looking at the cumulative log-return in the days surrounding a post, and ii) looking at the cumulative abnormal return. we follow the methodology of (Wan et al. 2021), which analyzes market movements around news events by looking at changes in abnormal return (AR). AR is derived from the Capital Asset Pricing Model (CAPM) (Fama & French 2004), which describes returns for company i at time t as follows:

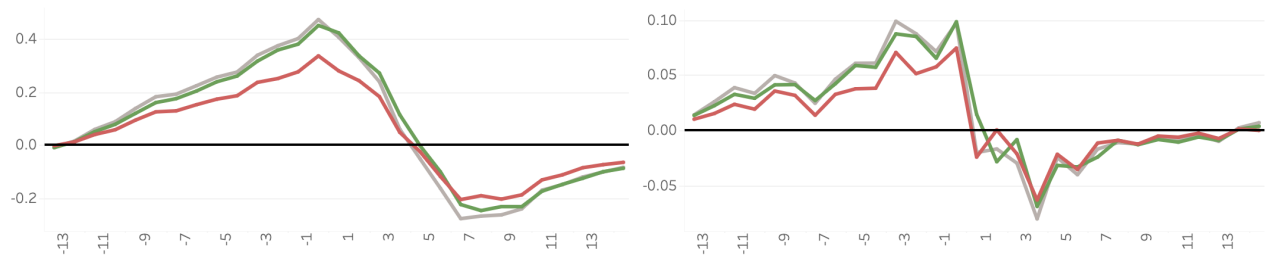
$$r_{i,t} = \alpha_t + \beta_t r_{m,t} + \epsilon_{i,t}. \tag{3.2}$$

Here $r_{i,t}$ is the log-return in the price of stock i on day t compared with the previous day, where $r_{i,t} = \log(\frac{p_{i,t}}{p_{i,t-1}})$ and $p_{i,t}$ is the adjusted close stock price. $r_{m,t}$ is the return of the market (in our case, we use the S&P 500), hence β captures stock price moves that are driven by movements in the wider market. $\alpha_{i,t}$ captures stock over/under-performance relative to the market. $\epsilon_{i,t}$ is a stochastic error term, often referred to as abnormal return (AR).

AR tends to have high magnitude in the presence of sudden price shocks (for example, a news story about a particular company), and hence it is often used for event detection in financial markets. In our case, it is useful for assessing if WSB sentiment is able to provide indication about a future price shock. We fit the CAPM model to each stock and day in our data, using a moving 180 day window. Following (Wan et al. 2021), we calculate the *seven-day cumulative abnormal return* (CAR) in stock i preceding day t:

$$CAR_{i,t} = \sum_{t-6}^{t} \epsilon_{i,t}. \tag{3.3}$$

Once we round the timestamp of a WSB submission to the nearest future market close time, we can match the CAR for each company/time to the WSB data, along with a time series of how the CAR changed over the 14 trading days preceding and following a WSB post.

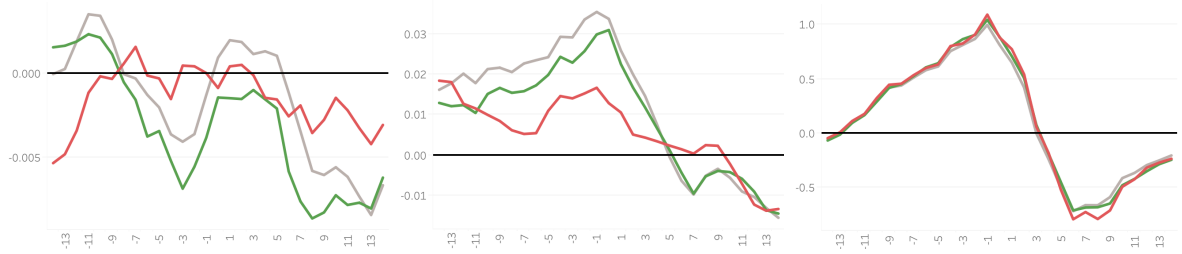


(a) Seven Day CAR

(b) Seven Day Cumulative Returns

Figure 3.5: Average seven-day cumulative return and CAR in 14 trading days around a submission, grouped by post sentiment (the red line displays the returns / CAR for bearish post sentiment, green line displays bullish, and grey line displays neutral).

CAR and Returns Results Figure 3.5 shows how the average seven-day cumulative abnormal return (CAR) and cumulative seven-day returns 14 trading days before and after a submission. Figure 3.5a shows the CAR plots for all WSB posts, while Figure 3.5b studies returns. We observe the following: firstly, CAR tends to be rising up to the submission date, and rapidly declines afterwards. This suggests that WSB activity tends to follow significant price changes in the market, rather than providing a leading indicator of price changes. Secondly, the shape of the curves are quite similar, regardless of sentiment breakdown – it is primarily the magnitude of the CAR that differs. This suggests that high CAR and returns are associated with more posts of any sentiment, although the effect is more pronounced for bullish posts.



(a) AAPL Seven Day CAR (b) TSLA Seven Day CAR (c) GME Seven Day CAR

Figure 3.6: Average seven-day CAR in 14 trading days around a submission for specific stocks, grouped by post sentiment (the red line displays the CAR for bearish post sentiment, green line displays bullish, and grey line displays neutral). CAR plots for twenty of the most popular stocks on WSB are presented in an interactive dashboard.

We present a breakdown of the CAR plots for twenty of the most popular stocks on WSB in an interactive dashboard.[8] The results show that the pattern presented in Figure 3.5a is most prevalent for 'meme' stocks with a low market capitalization. Stocks that have a broad following outside of the WallStreetBets forum, such as AAPL or MSFT, the CAR appears flat before and after submissions. Figure 3.6 displays an excerpt of these results. In an extension of the research in this thesis, we use the data to estimate the degree of trend-following and peer effects for different assets and observe that the two effects can have very

[8] https://sites.google.com/view/wsbtrialsite

different magnitudes for different assets within the forum.

An alternative set of experiments in Appendix 3.C underscores our findings that prices tend to run-up shortly before a post and decline after.

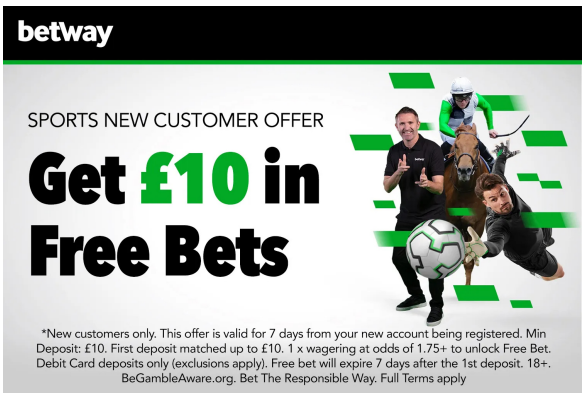
Appendix

3.A How prevalent are hype traders?

(a) Trading platform ad - example 1

(b) Trading platform ad - example 2

(c) Trading platform ad - example 3



(d) Gambling ad

Figure 3.A.1: **Sample trading and gambling advertisements**; We note the similarity in the style in certain trading advertisement to that of a gambling advertisement - encouraging the retail trader to get 'hooked' with an initial free share offering, or to get in with virtually no money in their account.

This study focuses primarily on the WSB discussion forum. However, a potential outstanding question is whether our findings extend to the broader trading population. We present several facts to support the broader relevance of our findings. Firstly, we note that anonymous stock market related forums have skyrocketed in their popularity. WSB, as we had noted previously, has experienced exponential growth and currently boasts fourteen million followers – putting these numbers into perspective, The Times newspaper recently

boasted 7.5 million subscriptions.[9] However, retail trader appetite for the hype is not satiated, as new forums are gaining popularity: r/StockMarketLeakz was one of the top-growing subreddit forums, while r/personalfinance, r/CryptoCurrency, r/bitcoin, r/stocks are all among the top 100 forums by number of subscribers, as of July 2023.

The rise of hype traders has not gone unnoticed as a prospective business opportunity with a tremendous rise in the number of retail trading platforms. Even though many cite long-term investment as a key reason to join, others use an advertisement approach with clear parallels to the gambling industry: offering a free trade or a free stock upon taking up the platform, or to trade with virtually no initial money in their account, as shown in Figure 3.A.1.[10,11,12,13] Other trading platforms, in turn, provide investors the opportunity to seamlessly execute upon the social and psychological biases studied within this text: 'Copy-Trader', for example, on the platform eToro allow investors to automatically copy the trades of others. A recent report by the UK Parliament discusses several important anecdotes pointing to the future relevance of this study: i) UK's largest investment platform, Hargreaves Lansdown, reported a 40% jump in net new business in the final six months of 2020, ii) the average age of platform users has dropped from 54 in 2012, to 47, reflecting a rise in younger investors (which are more prevalent on discussion forums, such as WSB), iii) Trading 212 (a different platform) announced on 2 February 2021 that it would temporarily pause new account openings due to huge demand.[14] Even though investor discussion forums have had an influence on markets in the past, evidence points to the fact that the popularity of WSB

[9] https://www.nytimes.com/2021/02/04/business/media/new-york-times-earnings.html
[10] https://www.wallstreetsurvivor.com/robinhood-free-stock/
[11] https://freetrade.io/legal/ps500-free-share-bundles-april-2022
[12] https://europe1.discourse-cdn.com/business20/uploads/trading212/original/1X/0d0133c0864441c22fd265d247422b8d5c37e5cb.jpeg
[13] https://talksport.com/sport/betting-tips/1304509/sports-betting-offer-premier-league-racing-free-bets-betway/
[14] https://commonslibrary.parliament.uk/the-rise-of-armchair-retail-trading-risks-and-regulation/

has permanently altered the composition and behaviour of traders[15] and has changed the dynamics of financial markets, as evidenced in this paper.

3.B Clusters and Returns

Author Clusters We observe eleven investor clusters, where assets are connected to other assets based on whether a similar subset of authors interacts in discussions about both assets simultaneously. In this section, we investigate the profitability of investors on WSB within these different clusters, as well as the correlation of asset returns within these clusters. We remind the reader of the most central assets within each of the discussion clusters: 0) DAL, UAL, 1) PLTR, 2) AMD, AAPL, MU, 3) SLV, GLD, 4) TSLA, SPY, 5) GME, AMC, 6) CCL, RCL, 7) LMT, BA, 8) JNUG, NUGT, 9) TLRY, CGC, 10) NKLA.

Table 3.B.1: **Distribution of Log-Returns for Investor Clusters**; we present some summary statistics for the next day log-returns for a portfolio invested according to the sentiment of submissions in different asset clusters on WSB as presented in our *Submission Network*. The table presents several summary statistics, as well as the p-value for the mean of the distribution to be equal to zero. Cluster **8** has insufficient observations and, therefore, is not considered.

	μ	σ	kurtosis	p-value	# posts	*Within Cluster Asset Correlation*
Cluster						
0	-0.0211	0.09	1.99	0.00	593	0.80
1	-0.0101	0.09	1.58	0.00	5,567	0.36
2	-0.0016	0.04	18.23	0.00	14,089	0.22
4	-0.0012	0.07	44.03	0.04	13,006	0.23
5	-0.0791	0.44	1.56	0.00	61,787	0.17
6	0.0062	0.10	5.99	0.21	398	0.84
7	0.0077	0.09	2.32	0.00	2,324	0.35
9	-0.0241	0.21	3.81	0.00	867	0.60
10	-0.0095	0.12	1.34	0.11	430	0.10

Table 3.B.1 presents our key result. We observe that portfolios built using opinions from WSB fairly consistently loose money. The tendency is particularly stark in the "meme" stock

[15] https://markets.businessinsider.com/news/stocks/reddit-retail-investor-trend-survey-social-media-wallstreetbets-wsb-stock-2021-4-1030366600

cluster – cluster **5**. This perhaps highlights the tendency for less sophisticated investors to participate in conversations surrounding these assets. A distinct, more profitable cluster appears to be cluster **7**. Unlike several of the other asset clusters, the assets appear to be more loosely connected, even though a potential unifying theme of large-scale engineering / space exploration is noticeable, since the cluster contains BA (Boeing), LMT (Lockheed Martin) and SPCE (Virgin Galactic Holdings). This analysis, in combination with our portfolio-building exercise with DD posts, highlights that overall investors on WSB appear to be trend-followers and fail to predict the market, however, small pockets of investors may have meaningful insights and outperform the market.

Table 3.B.2: **Distribution of Log-Returns for Topic Clusters**; we present some summary statistics for the next day log-returns for a portfolio invested according to the sentiment of submissions in different asset clusters on WSB. The Table 3.B.2 below presents several summary statistics, as well as the p-value for the mean of the distribution to be equal to zero. The clusters of GNUS (cluster 7), GTE (cluster 10), MGNI (cluster 12) and RC (cluster 17) are not considered as they do not meet a minimal threshold of 1000 posts.

	μ	σ	kurtosis	p-value	# posts	*# tickers in cluster*
cluster						
0	-0.0135	0.14	10.60	0.00	12,373	41
1	0.0027	0.03	4.33	0.00	2,530	1
2	-0.0211	0.13	9.57	0.00	3,312	44
3	-0.0301	0.18	2.71	0.00	9,898	12
4	-0.0168	0.13	21.04	0.00	6,557	17
5	-0.0005	0.06	30.90	0.41	10,045	72
6	-0.0018	0.03	17.60	0.03	1,580	1
8	-0.0029	0.06	58.04	0.00	5,763	35
9	-0.0639	0.37	0.57	0.00	19,686	1
11	-0.0742	0.16	3.63	0.00	11,349	33
13	0.0023	0.02	4.06	0.00	1,849	2
14	0.0130	0.09	1.60	0.00	2,160	2
15	-0.0095	0.13	10.90	0.00	2,474	13
16	0.0026	0.03	2.69	0.00	1,175	1
18	-0.0016	0.06	3.13	0.40	1,001	12
19	-0.0009	0.07	3.41	0.03	28,438	30

Topic Clusters Asset clusters identified by our topic graph isolate either individual tickers with highly distinct topics in their discussions, or broader groups of clusters with less dis-

tinct and more widely shared discursive environments; ticker topic clusters are displayed in Table 3.2. Table 3.B.2 assesses the statistical properties and behaviour of these asset clusters.

We observe, that most of the identified clusters with statistically significant p-values at the 1% threshold loose money. This suggests that on the whole, as reported above, strategies identified by mining the topic graph loose money. However, similarly to Table 3.B.1, certain clusters topic clusters exhibit positive daily returns, on average: cluster 1 (AAPL), 13 containing Microsoft and Morgan-Stanley, 14 containing NIO and XPEV two electric carmakers, and cluster 16 containing Alibaba. We notice that the smaller clusters with fewer assets perform better than much larger clusters, as they are the only ones displaying positive returns on average. On one hand, this may be driven by the systematic over-performance of tech stocks such as AMC, APPL and MSFT over the period. However, not all clusters with positive returns identified by the topic graph are composed of tech stocks, and conversely, not all the small clusters have positive returns. This allows us to suggest that tighter, more focused topic environments may indicate the presence of specialized 'pockets' of investors with better understanding and insights into the market.

3.C Asset returns around posting activity

We run an additional set of experiments testing the relationship between returns and sentiments (this experiment is done on the shorter dataset ending before the GME short squeeze). We average the sentiment characteristics in Eq. 3.1 by stock j and trading day t, denoting these mean sentiments by $\bar{\Phi}_{j,t}$ – where we are averaging individual author sentiments $\Phi_{i,j,t}$ as per Eq. 3.1. We merge these daily sentiment observations with US common stock returns reported by CRSP (detailed in Appendix 4.A), and transform the reported returns into log returns.

We first consider a regression of log returns on mean daily sentiment,

$$r_{j,t} = \lambda_{1,T}\bar{\Phi}_{j,t+T} + \eta_t^r + v_{j,t}^r, \tag{3.4}$$

where $v_{j,t}^r$ is a residual, η_t^r is a daily fixed effect, and T denotes a lag varying from -10 to 15 days. The OLS estimates for coefficients $\lambda_{1,T}$ describe how WSB sentiments are temporally related to stock returns.

Subsequently, we regress daily log returns on current mean sentiments, as well as previous day sentiments:

$$r_{j,t} = \lambda_1\bar{\Phi}_{j,t} + \lambda_2\bar{\Phi}_{j,t-1} + \eta_t^r + v_{j,t}^r, \tag{3.5}$$

where η_t^r is a daily fixed effect, $v_{j,t}^r$ an error term, and λ_1, λ_2 our coefficients of interest. This specification gives a sense for the dynamic properties of WSB sentiments – leading up to a trade day, plus their response on that day. In this analysis, we multiply log-returns by 100, for scaling purposes.

Results Figure 3.C.1 plots the OLS estimates for $\lambda_{1,T}$ in Eq. 3.4 as a function of lag T. Generally, past sentiments appear negatively related with current returns, although the effect is small, and not highly significant beyond four lags. Current sentiments are strongly correlated with current and past returns, and this effect is significant for up to five days in the past, before dissipating. This implies that a large return in an asset today will have a persistent impact on investor sentiment for five days into the future. Investor sentiments do not anticipate future returns, but rather follow the trend.

Table 3.C.1 reports OLS estimates for coefficients from Eq. 3.5 in Column 1. Returns again relate positively to contemporaneous sentiment, and negatively with previous day

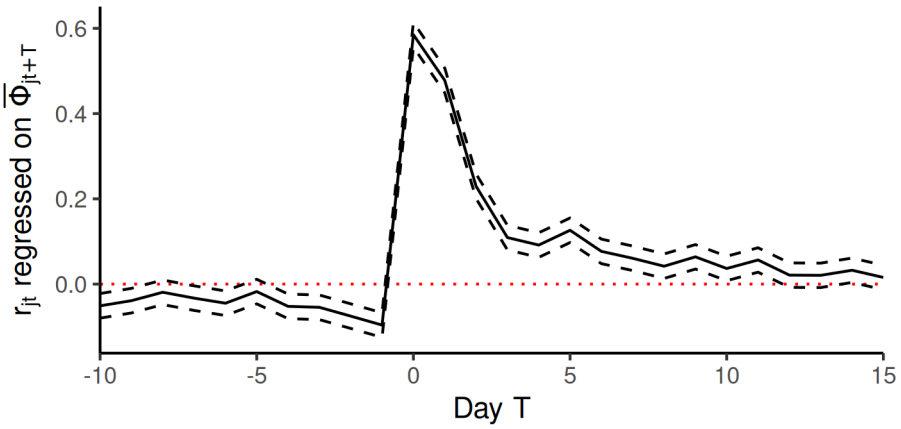


Figure 3.C.1: **Returns correlate negatively with past WSB sentiments, and positively with current and future WSB sentiments;** Daily stock log returns are regressed on daily average sentiments expressed in submissions on that stock on WSB, where sentiments are lagged by -10 to 15 days – the estimated relationship is described in Eq. 3.4. A lag of -5 implies that sentiments precede the returns observation by five days. The point estimate for the correlation is plotted by the solid black line, with the 99% confidence interval in dashed black lines. A correlation of zero is highlighted by the dotted red line. All variables are demeaned by their daily average. Return correlation with past WSB sentiments is negative, but strongly positive with current and future sentiments. Log-returns are multiplied by 100.

Table 3.C.1: Stock returns versus WSB characteristics

	Dependent variable:	
	$r_{j,t}$	
	(1)	(2)
$\bar{\Phi}_{j,t}$	0.60 *(0.04)* ***	
$\bar{\Phi}_{j,t-1}$	−0.16 *(0.02)* ***	−0.07 *(0.02)* ***
$r_{j,t-1}$		−0.06 *(0.004)* ***
$\bar{\Phi}_{j,t-1} \times r_{j,t-1}$		0.01 *(0.01)*
Day FE	Yes	Yes
Observations	8,287,639	8,287,639
R^2	0.0004	0.003

Notes: This table presents the OLS estimates for the relationship between stock log returns, $r_{j,t}$ and average expressed sentiment on WSB, $\bar{\Phi}_{j,t}$ and $\bar{\Phi}_{j,t-1}$. t represents time in days. The regression highlights the existence of a positive relationship between current sentiments and current returns, and a negative one between current returns and past sentiments. The negative relationship persists when controlling for previous day returns $r_{j,t-1}$ in Column (2). Accompanying standard errors, displayed in brackets, are clustered at the stock level, and calculated in the manner of MacKinnon & White (1985). Log-returns are multiplied by 100.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

sentiments. Both of these are statistically significant at the 1% level. However, the implied effects are relatively small; on an average day, returns are 0.1 log points lower if sentiments expressed on the previous day are twice more likely to be bullish than bearish.

In Column 2, we estimate Eq. 3.5 with the interaction between lagged, average sentiment

and lagged returns, to capture a non-linearity for sentiments in stocks that garner exceedingly high amounts of attention. The slight negative relationship between current and past returns could potentially confound the effect of past sentiment, as seen in the smaller coefficient for lagged sentiments. However, there is no clear evidence that the interaction between sentiments and outsized returns produce a significant effect on subsequent returns.

Chapter 4: Social Dynamics and Mechanisms of Sentiment Formation

What impacts the sentiments of investors active on the WSB forum? This chapter provides empirical evidence for the existence of two mechanisms – namely peer effects and extrapolation – among investors on WSB. The chapter that follows explores how these effects can impact asset prices by extending several models to incorporate these effects.

Estimating equation The target independent variable of interest for studying hype investor sentiment is the log-odds of bullish over bearish sentiment, modelled as

$$\Phi_{i,j,t} = g(b_{i,j,t}) + f(\bar{\phi}_{-i,j,(t-1,t)}) + \epsilon_{i,j,t}. \tag{4.1}$$

We remind the reader that $\Phi_{i,j,t}$ is our log-odds of bullish versus bearish sentiment expressed by user i about asset j at time t from our supervised ML labeling approach – Eq. 3.1 in Chapter 3.

An author i chooses a bullish over bearish strategy in asset j depending on: i) a signal $b_{i,j,t}$, and ii) the observed sentiments of peers, $\bar{\phi}_{-i,j,(t-1,t)}$. The goal of this chapter is to better understand $g(\cdot)$ and $f(\cdot)$. We set out to empirically test: whether $f(\cdot)$ is increasing (indicating peer effects), the importance of extrapolation in $g(\cdot)$, as well as other potential impacts documented in cognitive finance (such as reinforcement and surprise). After studying the mechanisms behind sentiment formation in this chapter, we consider potential implications for financial stability by developing several models for asset prices in Chapter 5 and empirically test market impact in Chapter 6.

Empirical strategy: consensus formation among investors A key challenge for estimating Eq. 4.1 is the existence of confounding variables, especially when estimating $f(\cdot)$ – peer

influence. We, therefore, use two approaches to estimate Eq. 4.1: i) the *Frequent Posters* approach, and ii) the *Commenter Network* approach. Both leverage different features of our data. For the *Frequent Posters* approach, we leverage the fact that certain users post multiple submissions about the same asset (hence, *frequent*) and the quasi-random variation in exposure to content on Reddit. For the *Commenter Network* approach, we use instances in which users comment on others' submissions in combination with an IV proposed in the networks literature (Zenou 2016, Patacchini & Zenou 2016, Bifulco et al. 2011) to more precisely gauge the transmission of sentiments about the same asset.

Consuming content on WSB Understanding the way in which content is consumed by users on WSB underpins our IV strategies. The options for sorting and filtering content on WSB are *identical* across users – put differently, two users who log on at the same time and choose the same sorting option will be exposed to identical content, regardless of their own posting or commenting history. Reddit does not have friendship/follower ties within specific forums.[1] Individuals cannot filter exposure to certain sentiments over others and do not receive specific, tailored notifications to view content related to their past commenting or posting history. To illustrate content consumption, consider two users **A** and **B**: **A** expressed bullish sentiments about AAPL, while **B** created several bearish posts (or commented on several bearish posts) about the same asset. If both users were to subsequently view WSB simultaneously (within our sample period), they would be exposed to the same exact content (verified through Reddit documentation and website AB testing). The posts that an individual is exposed to on WSB depend on what other anonymous, disconnected users have posted on the forum shortly before the author logs on, and what topic has recently gained popularity. This exposure of users to different opinions (similar in spirit to Weidmann & Deming (2021)) allows us to estimate direct peer effects.

[1] Followership ties on Reddit more broadly and the fact that they do not impact our approach are discussed in Appendix 4.B.5.

4.1 Identifying peer influence: Frequent Posters

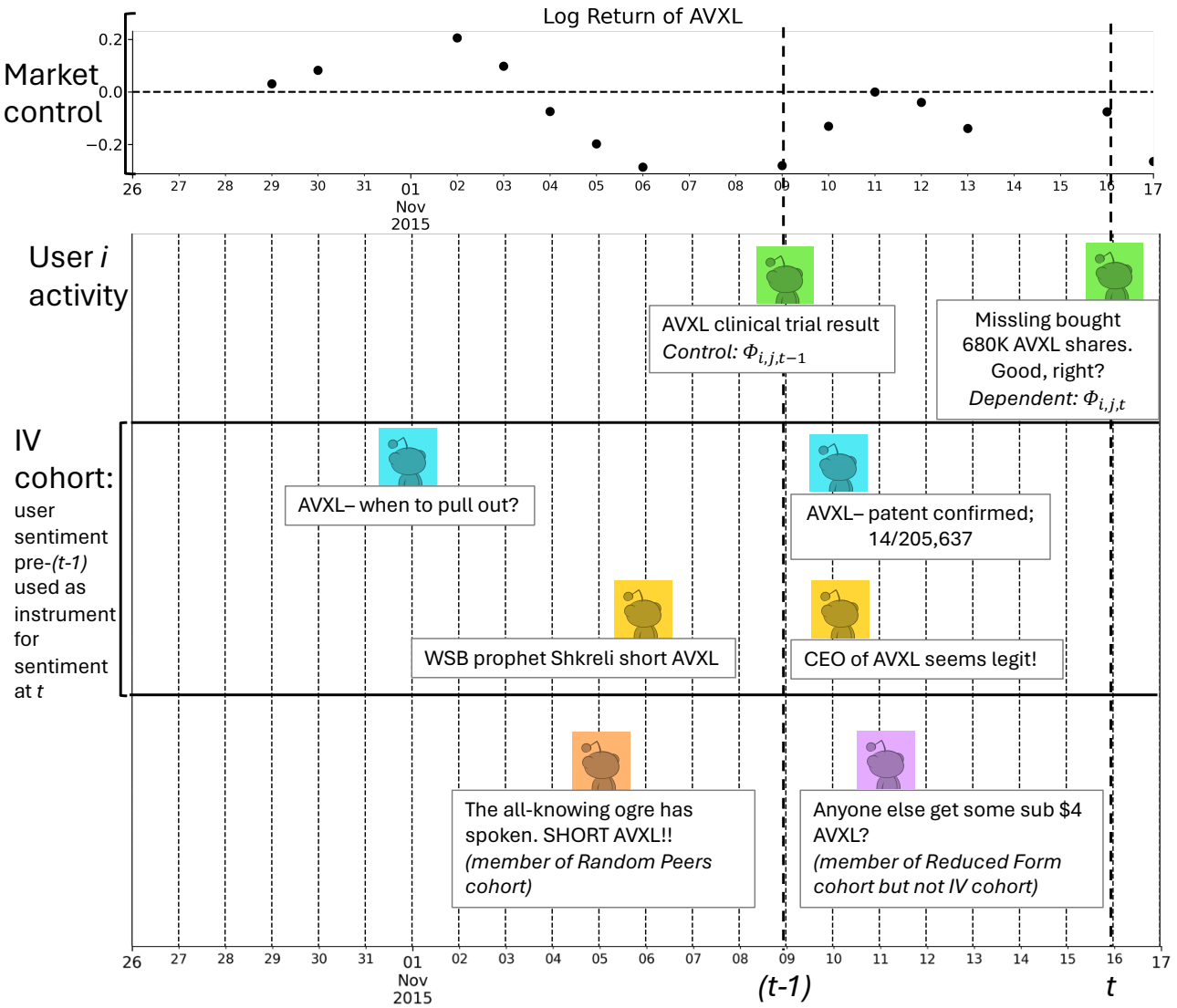


Figure 4.1: **Illustration of Frequent Posters Approach using Observations from WSB Data**; we observe user i create two posts about asset j (AVXL). She observes posts of peers between her two posts at time $(t-1)$ and t. In our IV approach, we use previous sentiments of peers to predict their sentiments between times $(t-1)$ and t – leveraging opinions that are formed exogenously to market conditions between $(t-1)$ and t. Market returns, as well as ticker fixed effects, are included as a control.

For the *Frequent Posters* approach, we observe that a subset of authors create at least two submissions about the same ticker. We quantify peer influence by identifying the impact of other authors who write submissions about the same asset *between* an individual's two submissions. We use an IV of previous, expressed peer sentiments to control for exogenous shocks (see Figure 4.1 for an illustration and Appendix 4.B.3 for a Directed Acyclic Graph (DAG) illustration). Our approach allows us to control for the author's sentiment prior to

exposure to his peers, in addition to market moves. We complement this approach with the *Commenter Network* approach, described in the next section.

Within WSB, we observe author i initially express a sentiment about an asset j, $\Phi_{i,j,(t-1)}$ (the continuous log-odds of a post expressing bullish over bearish sentiment, as per Eqs. 3.1&4.1), and, subsequently, write a new submission about the same asset at a later time, with an updated sentiment $\Phi_{i,j,t}$ (where time t is in event time). In the time between these posts, the author may observe submissions by others on the same asset expressing average sentiment $\bar{\Phi}_{-i,j,(t-1,t)}$, in addition to outside information related to the asset. Our goal is to identify the effect that expressed peer sentiments have on changing author i's sentiment.

Reduced form We first estimate the effect of average peer sentiment between an author's two submissions with the following linear model:

$$\Phi_{i,j,t} = \kappa_1 \bar{\Phi}_{-i,j,(t-1,t)} + \kappa_2 \Phi_{i,j,t-1} + \kappa_3 r_{j,t} + \kappa_4 \bar{r}_{j,t} + \kappa_5 \sigma^2_{j,t} + \kappa_6 X_j + \epsilon_{i,j,t}, \tag{4.2}$$

where the controls include: stock-specific fixed effects X_j, author i's past sentiment $\Phi_{i,j,t-1}$, stock log returns on the day of the author's post $r_{j,t}$ and the five days prior $\bar{r}_{j,t}$, the variance in log returns on the five days prior to the author's post $\sigma^2_{j,t}$. Even though an individual cannot filter exposure to certain sentiments over others, an exogenous shock in the period $(t-1, t)$ may affect the views of both peers and the author in question simultaneously. For this reason, the OLS estimates do not enable us to precisely estimate peer influence.

Instrumenting peer sentiment To tackle this issue, we use the historical views of peers as an IV for their views expressed within $(t-1, t)$. Our choice of IV is founded in psychology: Ross et al. (1975) find that 'once formed, impressions are remarkably persevering and unresponsive to new input', with later studies, such as Anderson et al. (1980), supporting these findings.

We reason about our choice of IV through the example in Figure 4.1, which presents a true example from the WSB discourse. The author of interest i creates two posts about AVXL and observes the posts of peers in-between. We use the historical sentiments of peers (created before $t-1$) as our IV. In this way, we isolate variation in peer sentiment *exogenous* to market conditions and news about asset j between time $(t-1, t)$.

We perform several robustness checks to ensure the validity of our instrument (discussed in greater detail after our results): we show that *predicted* peer sentiment is not correlated with asset returns between time $(t-1, t)$ demonstrating that there is no selection bias for a subset of peers whose past sentiment correlates with future returns or news. This is further reinforced by us including several controls for returns. We check that our estimation is not a lagged news effect by replacing the peers with a random peer cohort which posts before time $t-1$, and find that the random peer cohort has no effect. We test that there is no selection bias for peers to post a second time only when their sentiment agrees with average forum sentiment about the stock and show that probability of posting is uniform with respect to the level of agreement with peers. We also elaborate our reasoning about our choice of IV using a DAG in Appendix 4.B.3.

We estimate user k's sentiment – a peer of investor i – about asset j, $\Phi_{k,j,t}$, based on the sentiment they expressed previously, $\Phi_{k,j,t-1}$, and control for asset returns at the time of their original post, $r_{j,t-1}$:

$$\Phi_{k,j,t} = \kappa_1^0 \Phi_{k,j,t-1} + \kappa_2^0 r_{j,t-1} + \epsilon_{k,j,t}^0, \tag{4.3}$$

where $\epsilon_{k,j,t}^0$ is an idiosyncratic error. The coefficient κ_1^0 estimates the true effect of an individual's historical sentiment. Controlling for $r_{j,t-1}$ allows us to accurately estimate κ_1^0, while controlling for confounders. Eq. 4.3 is estimated using a sample containing submissions by all authors who post multiple times. The F-statistic for this first stage estimate,

presented in Table 4.2, suggests that this is a strong instrument. Our choice of IV gives a good approximation for author sentiment, while allowing us to control for common shocks affecting the sentiments of peers and investor i in the period $(t-1, t)$. We use the predicted outlook of peers between an author's posts, $\hat{\Phi}_{-i,j,(t-1,t)}$, to estimate peer effects as our Second Stage regression, while keeping all other controls the same – historic peer sentiment is used for prediction. Appendix 4.B provides further details on our variable construction and method. Appendix 4.A describes the construction of market variables, and their matching to WSB data.

4.1.1 Results

Table 4.1 presents the results of our estimation with column (1) presenting OLS estimates for κ (from Eq. 4.2) using observed variation in peer sentiments, and column (2) using predicted variation in peer sentiments – independent variables are normalized with respect to their mean and standard deviation (explained further in Appendix 4.B). The *Frequent Posters* approach indicates that peer effects are approximately 1.5 times *more important* in individual sentiment formation, as compared to extrapolation. Our non-normalized coefficient estimates in Appendix 4.B in Table 4.B.2 of 0.19 on predicted peer sentiments means that doubling in the odds of peers expressing bullish over bearish sentiments increases the odds of a given submission to be bullish, over bearish, by 14.1%. In all cases, the robust standard errors, clustered at the ticker level, produce estimates statistically significant at the 1% level. First Stage estimates are presented in Table 4.2.

4.1.2 Support for identification – Frequent posters

One potential concern is that individuals who post multiple times about the same asset may differ from the rest of the population on the forum. If this were the case, our findings would not allow us to draw valid conclusions about the overall population of investors. We provide

Table 4.1: Peer influence: *Frequent Posters* – full regression estimates

		Dependent Variable: $\Phi_{i,j,t}$		
		Reduced Form (1)	Full Second Stage (2)	Random Peers (3)
Independent Variables	$\Phi_{i,j,t-1}$	0.154 (0.010) ***	0.129 (0.011) ***	0.158 (0.010) ***
	$\bar{\Phi}_{-i,j,(t-1,t)}$	0.055 (0.011) ***	0.036 (0.010) ***	0.005 (0.009)
	$r_{j,t}$	0.022 (0.004) ***	0.025 (0.005) ***	0.023 (0.004) ***
	$\bar{r}_{j,t}$	0.007 (0.004)	0.007 (0.005)	0.006 (0.004)
	$\sigma^2_{j,t}$	-0.003 (0.004)	0.003 (0.008)	-0.003 (0.004)
	Ticker Fixed Effects	Yes	Yes	Yes
	No. Observations:	14,396	11,122	14,371
	R^2:	0.12	0.08	0.11
	R^2_{adj}:	0.08	0.06	0.08

Notes: The dependent variable is individual investor sentiment about an asset, scaled continuously between $(-\infty, \infty)$, is estimated by the individual's previously expressed sentiment about the same asset ($\Phi_{i,j,t-1}$) and a set of market control variables ($r_{j,t}, \bar{r}_{j,t}, \sigma^2_{j,t}$), using OLS. The sentiment of peers ($\bar{\Phi}_{-i,j,(t-1,t)}$) is estimated in several ways. In Column (1), we use observed, average sentiment of peers between an author's two posts. In Column (2), we estimate the sentiment of peers using an IV. In Column (3), we select a random cohort to estimate peer sentiment. Robust standard errors, clustered at the ticker level, are presented in parentheses. Observations with incomplete market data are dropped.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Table 4.2: Peer influence in WSB sentiments – first stage

	Dependent Variable: Sentiment of Peers	
	Frequent Posters (1)	Network (2)
Historical Sentiment of Peers	0.31 *(0.01)* ***	
Sentiment of Neighbours' Neighbours		0.14 *(0.01)* ***
Author controls ($X^0_{i,j,t}$)	No	Yes
Controls for returns ($r_{j,t-1}$)	Yes	No
Number of obs.	19,370	24,013
F-statistic	1,105	118

Notes: This table presents the First Stage estimates for peer influence on WSB. In column (1), the First Stage is estimated using the initial sentiment expressed by an author about an asset to estimate his sentiment in the following post. In column (2), the First Stage is estimated using the sentiment of previous submissions that an author commented on, regarding the same asset – the second IV of the neighbor's own historical sentiment is presented in Appendix 4.B.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

evidence that sentiments expressed in our samples are similarly distributed to those of the overall user population in Appendix 4.B.

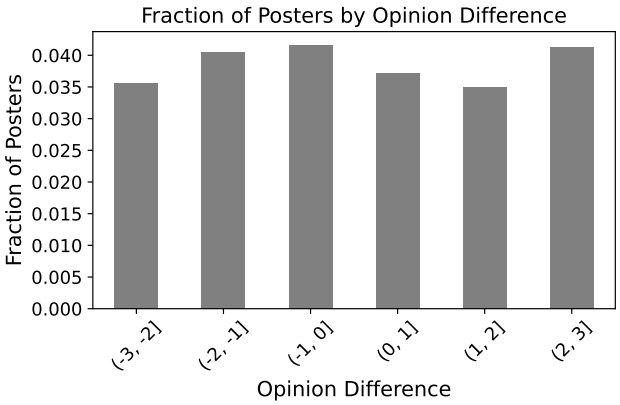


Figure 4.2: **Probability of posting on 5^{th} day bucketed by disagreement level with peers**; the fraction of individuals that post five days after their original post bucketed by the difference between their sentiment and that of peers (the difference is taken between the mean peer sentiment in the intermediate five days and the sentiment expressed by the author in their follow-up post).

Endogenous choice to post multiple times An additional concern may arise around the endogenous choice for certain individuals to post multiple times. Specifically, individuals may choose to post a second time: i) only if their peers agree with them, or ii) when their predictions about the market are correct (in which case the IV would be documenting author i's response to general market moves, rather than peer influence).

We test for (i) by looking at the probability that an individual posts X days after their original post depending on the difference between their sentiment and that of peers (the difference is taken between the mean peer sentiment in the intervening X days and the sentiment expressed by the author in their follow-up post). Figure 4.2 presents our results at the five day time horizon – we observe that the distribution for the probability of posting with sentiment difference is uniform, invalidating the idea that individuals who post are only those who peers agree with. Appendix 4.B presents the same results at different time horizons.

We test for (ii) by including the average return of stock j in the period between $(t-1)$ and t as an additional control. If individual i was swayed by market moves that coincide with the sentiment of peers, rather than peer effects themselves, this control should erode the significance of peer effects. However, adding this control does not change our results

and is not significant - this implies that it is not peers echoing market returns that sways sentiments, but rather the peer effects themselves that impact sentiment. Additionally, we test to see if our predicted peer sentiment, $\hat{\Phi}_{-i,j,(t-1,t)}$, is correlated with the returns on day t when user i posts – the two do not have a statistically significant relationship. Appendix 4.B further discusses both robustness checks.

Distinguishing information from peer effects Another concern is whether our proposed independent variables – asset price movements, ticker fixed effects and author historical sentiments – are effective controls for unobserved ticker characteristics. Along the same lines, one might worry about our ability to distinguish between peer effects and information being absorbed slowly from peers from their initial post before time $(t-1)$. If our controls in the *Frequent Posters* formulation and identification technique are valid, then a randomly selected cohort of individuals who post on the same ticker *before* the author's first post, should have no effect on the sentiments expressed in dependent submissions. This would demonstrate that what we observe is not information being observed slowly by peers or lingering forum sentiment, but rather peer effects in the absence of any additional information. The results are presented in Column (3) in Table 4.1. We find no statistically significant correlation from the randomly selected cohort, which rules out longer lags in market conditions manifesting as peer effects in user sentiments. We also run several robustness checks validating that peer effects persist if we select a cohort of peers that posts initially over three days before the author's post $(t-1)$. We observe that peer effects persist in this truncated sample, further invalidating the slow spread of information argument in favour of our peer effects estimation.

Longer time horizon – GME short squeeze and beyond A final question is whether the effects we study persist outside of our selected timeframe. We test our approach on the

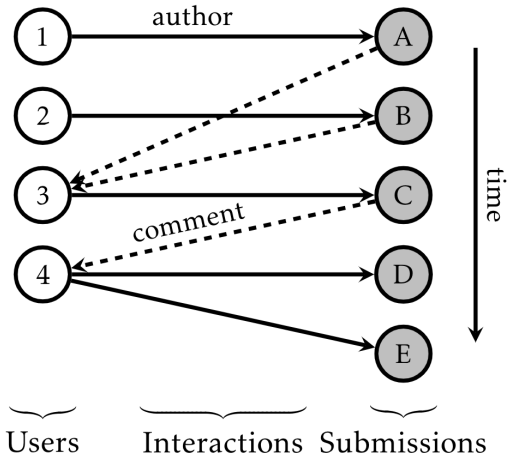
full dataset described in the previous section, which offers a time horizon spanning through mid-2021. We observe that through the end of the GameStop short squeeze, our estimate for peer effects is close to that of this chapter. However, after late February 2021, peer effects and extrapolation are no longer appear significant. The individual's own perspective about the stock, $\Phi_{i,j,t-1}$, not only remains significant but its coefficient also increases in magnitude. Full results are presented in Appendix 4.B. This is consistent with the findings of Bradley et al. (2024) documenting that the informational content and quality of the forum steeply declined after GME. Our analysis points to the fact that after GME, individuals ceased to update their perspectives based on that of their peers or the market, but became firmly entrenched in their own beliefs. This upholds the narrative that WSB transformed from an investor discussion platform into a social movement.

4.2 Identifying peer influence – Commenter Network

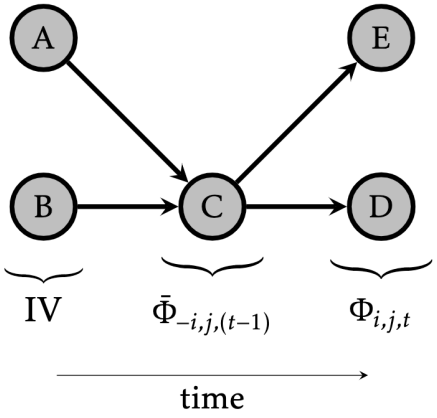
The *Commenter Network* approach considers a submission-to-submission network, with an earlier submission exerting peer influence on a future submission if the author of the later submission commented on the earlier one. The submission-to-submission network helps identify peers an author interacts with more precisely. Here, we also control for market moves, and employ a set of IVs to address endogeneity concerns. As our IVs, we measure: i) sentiments of submissions to which the influencing submission is connected (the 'friends of friends' – detailed in Figure 4.3b), and ii) the historic sentiment of neighbours. The underlying argument rests on the premise that neighbours of network distance two exert an influence on user sentiments through peer effects (consistently with Bond et al. (2012)). A user's endogenous choice to comment on certain posts over others would therefore not account for users one step removed.

WSB allows us to trace the interactions of users through a commenting network, even

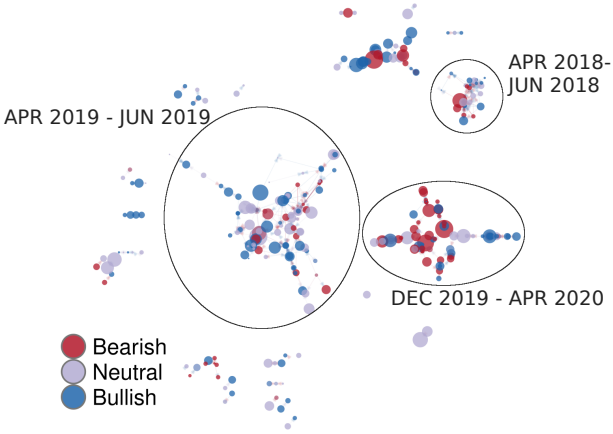
though the forum does not support friendship ties. We exploit a submission-to-submission interaction network for each asset, tracking which submissions in the past influence future submissions based on authors' commenting histories. This method offers a more precise way to identify a user's peers by observing which individuals, and submissions, an author explicitly interacts with. Figures 4.3a and 4.3b illustrate the approach.



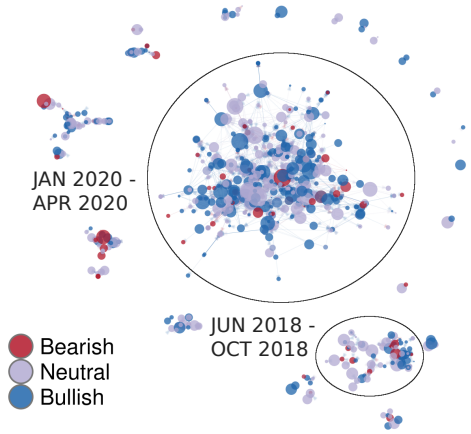
(a) Bipartite network between authors and submissions



(b) Submission-to-submission projection of network in Figure 4.3a



(c) Submission-to-submission network: DIS



(d) Submission-to-submission network: MSFT

Figure 4.3: **User networks in WSB conversations**; WSB data is summarised as a bipartite graph, illustrated in Figure 4.3a, where users (left) are linked to submissions (right) when they author the submission (solid edge) or comment on the submission (dashed edge). The resulting projection of submissions, in Figure 4.3b, tracks the propagation of sentiments Φ. The submission-to-submission networks for two stocks in Figures 4.3c and 4.3d reveal that individuals post more submissions that are bullish(bearish) at times when the price of an asset increases(decreases) dramatically, with some visual evidence that similar sentiments tend to cluster.

Two examples of submission-to-submission networks in our data are displayed in Figures 4.3c and 4.3d. Distinct temporal clusters emerge, as a certain asset gains and loses

prominence on WSB. Some discussions appear fragmented: the *DIS* discussion in Figure 4.3c, for example, contains several smaller clusters, with perceptible differences in overall sentiments. Others, such as the *MSFT* discussion in Figure 4.3d, contain a giant component where investors with different sentiments interact.

Our network approach uses a similar Reduced Form and Second Stage to the *Frequent Posters* approach in Eq. 4.2. We modify our control for an author's past sentiment about the stock to account for authors who post for the first time: a dummy variable encodes whether the author's most recent previous post is bearish, neutral, bullish or missing.

Instrumenting peer sentiment We use an IV approach to estimate peer influence. As the First Stage, we estimate the sentiments of neighbours to estimate an author's view. As indicated in Figure 4.3b, the sentiments in submissions **A**, **B** can be used to predict that of submission **C**. The *predicted* sentiment of **C** can then, in turn, be used to predict the sentiments of **D** and **E**. This choice of IV is well-established in the networks literature (Zenou 2016, Patacchini & Zenou 2016, Bifulco et al. 2011), and helps control for the exogenous choice to comment on certain submissions and not others – the method further benefits from the fact that users are not alerted to the activity of specific other individuals. We also include the neighbour's own historical sentiment, as a set of categorical variables, as the second IV (similarly to the *Frequent Posters* approach).

Timing of observations We use the timings of events to mitigate the common shock problem for both our IVs: the neighbour's historical sentiment and the 'friends or friends' submissions. For the latter, we calculate the time period of influence for a given post, which ends when the last comment is made on a submission. This effectively marks the point when a particular submission ceases to be of interest to the WSB community. We filter for instances where the period of influence for a submission used as an IV for another submis-

sion ends before the new submission we are modeling is created. In practice, if submission **C** in Figure 4.3b occurs on July 1st at 2:31PM, the final comments on posts **A** and **B** must occur before, in order to ensure that our IV is not affected by a common shock. We also include an author's own, historical sentiment as an IV only if his previous submission occurs at least two business days before the current one.

The *Commenter Network* offers certain upsides, but also certain shortcomings, as compared to the *Frequent Posters* approach. The network method more precisely identifies the channels of influence between authors. However, the allocation of peers is no longer random, since the network structure is governed by a *choice* to comment.

4.2.1 Results

Table 4.3: Peer influence: *Commenter Network* – full regression estimates

		Dependent Variable – $\Phi_{i,j,t}$		
		Reduced Form (1)	Full Second Stage (2)	Random Network (3)
Independent Variables	$\phi^{-1}_{i,j,t-1}$	-0.226 (0.026) ***	-0.215 (0.021) ***	-0.342 (0.040) ***
	$\phi^{0}_{i,j,t-1}$	0.047 (0.023) **	0.034 (0.022)	0.073 (0.036) **
	$\phi^{+1}_{i,j,t-1}$	0.160 (0.028) ***	0.141 (0.031) ***	0.244 (0.042) ***
	$\bar{\Phi}_{-i,j,(t-1,t)}$	0.041 (0.009) ***	0.022 (0.009) **	0.009 (0.009)
	$r_{j,t}$	0.020 (0.003) ***	0.023 (0.006) ***	0.031 (0.005) ***
	$\bar{r}_{j,t}$	0.005 (0.003)	0.006 (0.006)	0.008 (0.005)
	$\sigma^2_{j,t}$	0.044 (0.289)	0.512 (0.508)	0.078 (0.449)
	Ticker Fixed Effects	Yes	Yes	Yes
	No. Observations:	24,902	16,514	25,220
	R^2:	0.09	0.07	0.09
	R^2_{adj}:	0.06	0.06	0.06

Notes: The dependent variable is individual investor sentiment about an asset expressed in a single submission, scaled continuously between $(-\infty, \infty)$. We estimate it using the individual's previously expressed sentiment about the same asset ($\phi_{i,j,t-1}$) as a categorical variable, with the author not having posted previously ($\phi^{NA}_{i,j,t-1}$) as the baseline. We control for a set of market control variables ($r_{j,t}, \bar{r}_{j,t}, \sigma^2_{j,t}$). The sentiment of posts that the author commented on previously ($\bar{\Phi}_{-i,j,t-1}$) is estimated several ways. In column (1), we present the estimate using the sentiment of posts the author previously commented on. In column (2), we use an IV to predict the sentiment of posts the author comments on. In column (3), we randomly rewire the network, connecting the author to a random set of posts about the same ticker. Robust standard errors, clustered at the ticker level, are presented in parentheses. Observations with incomplete market data are dropped.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

The results of the *Commenter Network* approach are consistent with those of the *Frequent Posters* approach, presented in Table 4.3 – column (1) presents OLS estimates for using average, observed peer sentiment, while column (2) uses predicted sentiment from our IV approach. In both cases, the robust standard errors, clustered at the ticker level, produce estimates statistically significant at the 1% level. First Stage estimates for estimating peer sentiment using their neighbors are presented in Table 4.2, while the estimates for using the neighbour's own historical sentiment are presented in Appendix 4.B.

The estimated coefficients in both approaches suggest that an exogenous increase in average peer outlook appears to increase an investor's own future view about an asset. These findings suggest that the data supports a model where strategic complementarities govern the investment decisions of retail traders sampled on WSB and where extrapolation plays an important role.

4.2.2 Support for identification – Commenter network

One potential concern is that individuals who comment on other's posts may differ from the rest of the population on the forum. If this were the case, our findings would not allow us to draw valid conclusions about the overall population of investors. We provide evidence that sentiments expressed by our sample are similarly distributed to those of the overall user population in Appendix 4.B.

This approach is more comprehensively discussed with the networks literature, so we rely on the established robustness checks. If our controls are useful in the *Commenter Network* formulation, a random rewiring of the network should yield no effect. The results are presented in column (3) of Table 4.3: no statistically significant correlation emerges from the randomly selected cohort. This provides further evidence that unobserved factors influencing within-ticker variation in both peer composition and author sentiment are not

confounding.

An additional concern with our *Commenter Network* approach is overidentifying restrictions. A J-statistic of 0.43, and a corresponding p-value of 51%, leads us to believe that our additional instruments are exogenous (see Appendix 4.B for further details).

4.2.3 Further insights

WSB data provide additional opportunities to test investor responses to a market surprise, and the reinforcement mechanism between peers and asset prices. We consider the sentiments expressed by investors i about asset j at time t, $\Phi_{i,j,t}$, as our dependent variable and use the controls from Eq. 4.2 to test for two additional effects: market surprise and reinforcement.

We define two types of surprises: i) a positive surprise if asset j experiences a return which is two standard deviations higher than the 30-day historical average for the stock on day t or on the day before, and ii) a negative surprise if asset j experiences a return which is two standard deviations lower than the average for the stock on day t or on the day before. We compute the average and standard deviation for stock j using data of the thirty trading days before t. We also interact returns and predicted peer sentiment to see the extent to which peer effects are reinforced by returns. We use the *predicted* peer sentiment from our *Frequent Posters* approach in our regressions to control for sentiments that respond to current price changes.

Surprise Table 4.4 presents the results from our exploration of surprise and reinforcement. Column (1) contains the OLS estimates when including positive and negative categorical variables for market surprise. A negative market surprise appears to significantly affect investor sentiments. The result is not symmetric – a positive surprise does not appear to convince investors of the upside potential of a stock. This observation suggests that

Table 4.4: Additional effects: surprise and reinforcement

		Dependent Variable: $\Phi_{i,j,t}$	
		(1)	(2)
Independent Variables	$\hat{\Phi}_{-i,j,(t-1,t)}$	0.037 *(0.009)* ***	0.036 *(0.012)* ***
	$r_{j,t}$	0.019 *(0.006)* ***	0.017 *(0.005)* ***
	Positive Surprise	-0.018 *(0.031)*	
	Negative Surprise	-0.114 *(0.038)* ***	
	$(r_{j,t} \times \hat{\Phi}_{-i,j,(t-1,t)})^{+}$		0.074 *(0.025)* ***
	$(r_{j,t} \times \hat{\Phi}_{-i,j,(t-1,t)})^{-}$		0.070 *(0.108)*
	Author & asset controls ($X_{i,j,t}$)	Yes	Yes
	No. Observations:	11,073	11,116
	R^2:	0.08	0.08
	R^2_{adj}:	0.06	0.06

Notes: The dependent variable – individual investor sentiment about an asset, scaled continuously between $(-\infty, \infty)$ – is estimated using the variables in Eq. 4.2 and additional variables, using OLS. The additional variables in column (1) are categorical variables for positive and negative market surprises at time t in asset j; in column (2) the additional variables are a cross term between asset j's returns and the estimated sentiments of peers: $(r_{j,t}$ X $\hat{\Phi}_{-i,j,(t-1,t)})^{+}$ is the product if the predicted sentiment of peers is positive and returns are also positive, and zero otherwise; $(r_{j,t}$ X $\hat{\Phi}_{-i,j,(t-1,t)})^{-}$ is product if the predicted sentiment of peers is negative and returns are also negative, and zero otherwise. $(r_{j,t}$ X $\hat{\Phi}_{-i,j,(t-1,t)})^{+}$ captures the extent to which positive peer predictions correspond to observed market moves; the reverse is true for $(r_{j,t}$ X $\hat{\Phi}_{-i,j,(t-1,t)})^{-}$. Peer sentiment $\hat{\Phi}_{-i,j,(t-1,t)}$ is estimated using the *Frequent Posters* approach to control for confounders. Robust standard errors, clustered at the ticker level, are presented in parentheses. Observations with incomplete data are dropped. *** Significant at 1% level ** Significant at 5% level * Significant at 10% level

downside panic spreads quickly within the investor population. This effect is in addition to the significant impact returns have on sentiment.

Reinforcement Column (2) considers the effect from market reinforcement of peer sentiments by including the cross term between returns and the predicted sentiments of peers. The cross terms are separated depending on whether the predicted peer sentiment $\hat{\Phi}_{-i,j,(t-1,t)}$ is positive or negative: $(r_{j,t} \times \hat{\Phi}_{-i,j,(t-1,t)})^{+}$ is the *bullish* interaction when $\hat{\Phi}_{-i,j,(t-1,t)}$ and returns are both positive and zero otherwise, whereas the bearish interaction $(r_{j,t} \times \hat{\Phi}_{-i,j,(t-1,t)})^{-}$ is positive if predicted sentiment and returns take negative values, zero otherwise. Therefore, a large value for the bullish interaction corresponds to peers forecasting positive returns in asset j and the asset j simultaneously experiencing positive returns on the day of author i's submission.

In Table 4.4, the bullish interaction is highly significant. WSB users are spurred by peers predicting positive returns and subsequently observing the asset outperform in the market, possibly suggesting 'irrational exuberance' (Shiller 2005). The reverse is not true for bearish reinforcement.

Appendix

4.A Market variables

We include a set of market return and volatility control variables. The data source for these variables are the daily stock files issued by the Center for Research in Security Prices (CRSP), accessed through Wharton Research Data Services.

The following market variables serve as controls.

$r_{j,t}$: the log return for asset j on trading day t. From CRSP, we calculate it using their 'RET' variable: $r_{j,t} = \log(RET_{j,t} - 1)$, which automatically corrects the percentage change in closing prices for share splits and dividend distributions.

$\bar{r}_{j,t}$: the average log returns for asset j in the five days prior to t (the log return on day t is not included). A minimum of three daily log-return observations is required, otherwise the observation is set as missing.

$\sigma^2_{j,t}$: the variance of log returns for asset j in the five days prior to t (the log return on day t is not included). A minimum of three daily log-return observations is required, otherwise the observation is set as missing.

Matching submission timings to trade timings If a post occurs before 16:00:00 EST on day t, we match it with the log-return on the same day t. If a post occurs after 16:00:00 EST on a given day, we match it with market data for the next trading day, $t + 1$. This is done to capture the fact that many news announcements occur after hours and someone posting after the market close may be exposed to these after-hour moves. Instance in which submissions are made on weekends, or holidays, are matched to the next possible trading day. For example, a submission made at 5pm on Friday is paired to the observed log return for the following Monday.

4.B Additional results and robustness checks for peer effects

4.B.1 First stage regression estimates

Tables 4.1 and 4.3 present our full regression estimates. Table 4.B.1 presents our First Stage estimates for our *Commenter Network* approach, which has multiple IVs.

Table 4.B.1: First Stage estimates for *Commenter Network* approach

	$\phi^{-1}_{i,j,t-1}$	$\phi^{0}_{i,j,t-1}$	$\phi^{+1}_{i,j,t-1}$	$\bar{\Phi}_{-i,j,t-1}$
Dependent variable:				
Sentiment of Peers	-0.30 (0.04) ***	0.12 (0.03) ***	0.25 (0.03) ***	0.14 (0.01) ***

Notes: The dependent variable is individual investor sentiment about an asset expressed in a single submission, scaled continuously between $(-\infty, \infty)$, modeled using IVs. We estimate it using the individual's previously expressed sentiment about the same asset ($\phi_{i,j,t-1}$) as a categorical variable, with the author not having posted previously ($\phi^{NA}_{i,j,t-1}$) as the baseline, as well as the average sentiment of posts that the author commented on previously ($\bar{\Phi}_{-i,j,t-1}$). We use the timing of IVs to control for common shocks, as discussed in the main text. Our regression has 24,013 observations and an F-statistic of 118.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

4.B.2 Normalization procedure and non-normalized coefficient estimates

In order to compare the relative impacts across variables, we perform mean / standard deviation normalization on the non-categorical variables within our regression. We normalize market variables with respect to the log-returns of all assets discussed on WSB since the forum's creation in 2012. We normalize sentiment variables with respect to the observations within our regressions. The normalization is performed in order to be able to compare the impact of peer effects and returns on sentiment formation.

Table 4.B.2 presents the non-normalized coefficient estimates for our second stage. We observe that the coefficient on returns and predicted peer sentiment are both higher, and the coefficient on returns is larger than that of predicted peer sentiment: these changes relate to the standard deviations of the two variables which are 0.037 for daily returns and 0.271 for predicted peer sentiment, *Frequent Posters*, and 0.168 for predicted peer sentiment, *Com-*

menter Network. The first phenomenon is explained by the fact that the standard deviation for both variables is less than one; the second is explained by the fact that daily returns have a substantially smaller standard deviation than that of peer sentiment.

Table 4.B.2: Peer influence in WSB sentiments

	Frequent Posters (1)	Network (2)
Second Stage – peer influence estimated using *predicted* average sentiment of peers *(non-normalized)*		
	Dependent Variable: Investor Sentiment ($\Phi_{i,j,t}$)	
Average peer sentiment, $\hat{\Phi}_{-i,j,(t-1)}$ *(predicted)*	0.198 (0.053) ***	0.197 (0.083) **
$r_{j,t}$	0.993 (0.185) ***	0.921 (0.251) ***
Author & asset controls ($X_{i,j,t}$)	Yes	Yes

Notes: this table presents the non-normalized coefficient estimates for the Second Stage of our *Frequent Posters* and *Commenter Network* regressions.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

4.B.3 Evidence of identification strategy – Frequent Posters

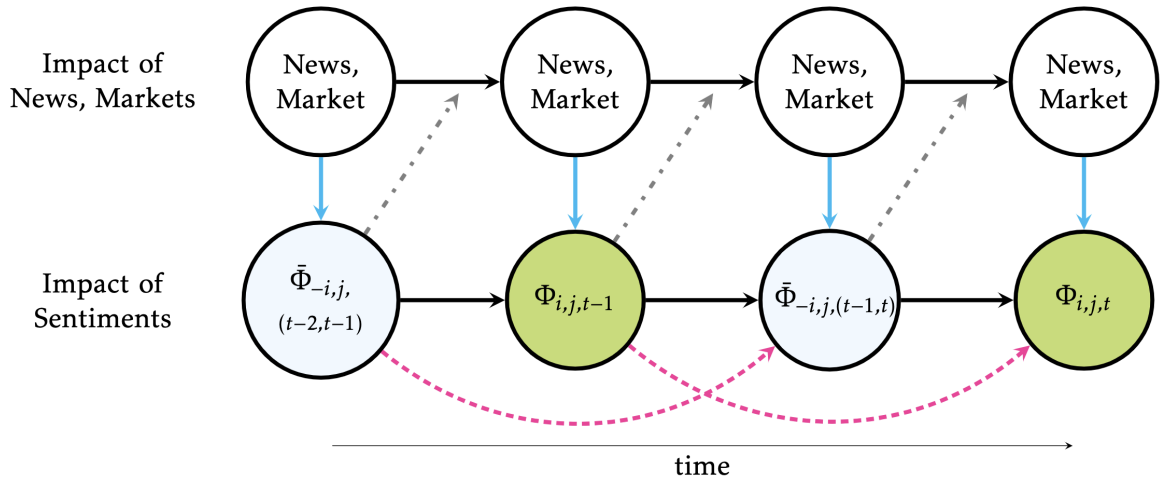
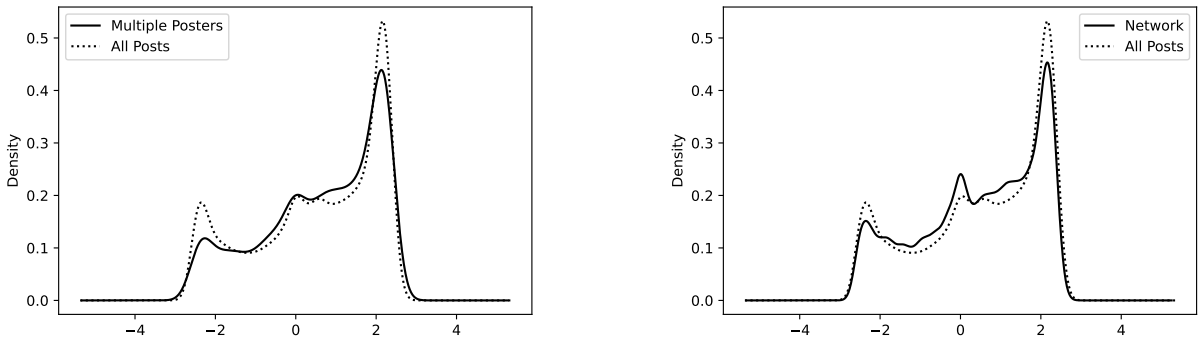


Figure 4.B.1: ***Frequent Posters* Directed Acyclic Graph (DAG);** we trace the flow of information within our system. Arrows represent the impact that information from one source has on the next source. Light blue nodes $\bar{\Phi}_{-i,j,(t-2,t-1)}$ and $\bar{\Phi}_{-i,j,(t-1,t)}$ represent peer sentiment; green nodes $\Phi_{i,j,t-1}$ and $\Phi_{i,j,t}$ represent the sentiment of investor i - our target variable. Time t is expressed in event time. The magenta, dashed line represents the impact that historical sentiment expressed about an asset has on the author's own future opinion due to people being reticent to chance their minds (Ross et al. 1975). In our first stage, we estimate $\bar{\Phi}_{-i,j,(t-1,t)}$ by $\bar{\Phi}_{-i,j,(t-2,t-1)}$, controlling for the market move at the time of the peer's initial post while estimating the coefficients. In this way, we are able to isolate the impact that peer sentiment $\bar{\Phi}_{-i,j,(t-1,t)}$ has on individual i at time t, $\Phi_{i,j,t}$.

Channels of Influence - Directed Acyclic Graph (DAG) We present the channels of influence used within the *Frequent Posters* framework in Figure 4.B.1. We reason about our

choice of IV through the Directed Acyclic Graph (DAG) shown in Figure 4.B.1. We consider that historic news and market moves are fully reflected in the news and market information available within the following timestep. Information shared by peers is also fully incorporated from one timestep to the next; however, dotted pink lines indicate the persistence of individual author sentiments (the persistence of individual formed impressions). Leveraging the structure of our DAG, we estimate investor k's sentiment (a peer of investor i) about asset j, $\Phi_{k,j,t}$, based on the sentiment they expressed previously, $\Phi_{k,j,t-1}$, and control for asset returns at the time of their original post,



(a) Frequent Posters Sentiment PDF

(b) Commenters Sentiment PDF

Figure 4.B.2: **Density Plot of Sentiments Expressed on WSB**; We present the density plot of the sentiments expressed by users on WSB who post multiple times, labeled as *Multiple Posters*, those who comment on others' posts, labeled as *Network*, and that of all submissions, labeled as *All Posts*.

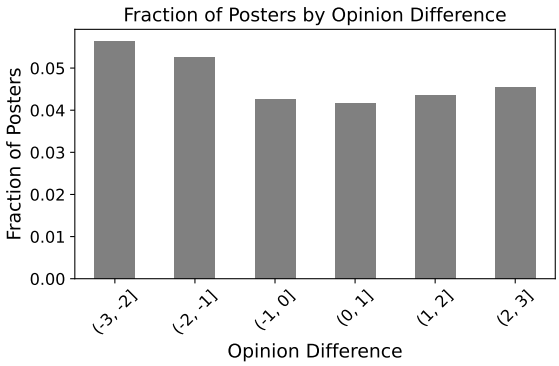
Distribution of sentiments A potential concern with our approach is whether the sentiments expressed by individuals who post multiple times follow the same distribution as all submissions on the forum. Figure 4.B.2a presents the distribution of sentiments for the second or later post of an author about a ticker, which provides evidence that the sentiment distributions are similar to that of other posters on WSB. This supports the hypothesis that our analysis offers insight into how all individuals on WSB form opinions.

Random peers A second concern is whether we effectively control for unobserved ticker characteristics. We replace the composition of an author's peers with a random cohort of

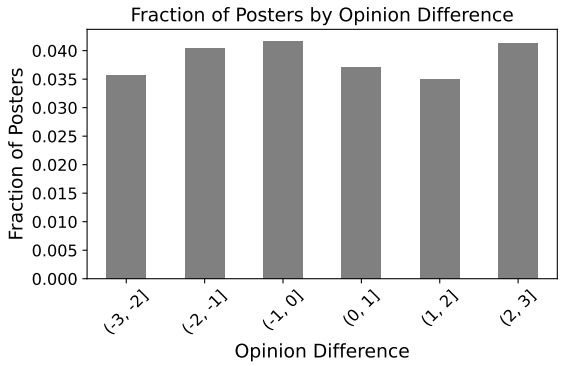
people who post on WSB about the same ticker. The random cohort is chosen as follows. We observe how many peers an individual author has. We then select a random sample of the same number of individuals, without replacement, who do not post between an author's two post but post about the ticker at a different time before $(t-1)$ for the *Frequent Posters* approach (if fewer individuals post before, we select all of those individuals). The results are presented in Tables 4.1, column (3). We observe that all the coefficients remain close to their original values, except for the peer effect, which becomes insignificant. This lends credibility to our peer identification strategy and shows that unobserved factors that influence within ticker variation are not confounding our estimates. This also supports our argument that we are not observing a longer time period of information absorption from peers – if this were the case, one would expect the random cohort of peers that post before time $(t-1)$ to remain significant.

Endogenous choice to post multiple times One potential concern is that individuals who post multiple times are those that likely agree with their peers, whereas the sentiments of those that disagree will drop out of the forum and not create a second post. We, therefore, look at the probability of individuals creating a follow up post at various time horizons, depending on the distance between their own sentiments about an asset and that of peers in the interim period. Figure 4.B.3 shows the probability of a user creating a follow up post, plotted in different buckets depending on the magnitude of disagreement between them and their peers – we select several time horizons between four and seven days when an author is most likely to create a follow-up post. We observe that individuals are *not* more likely to create posts if they agree with peers – if this were the case, we would expect higher probabilities of creating a follow-up post in the $(-1,0]$ and $(0,1]$ buckets (where the sentiment difference between the individual and the average sentiment of peers is lower). We, therefore, observe that differences in opinion persist among individuals who post multiple

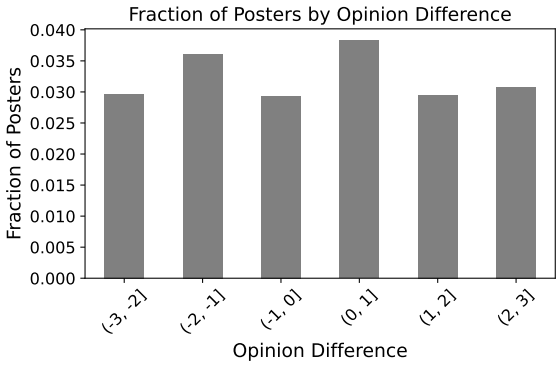
times about the same asset on WSB.



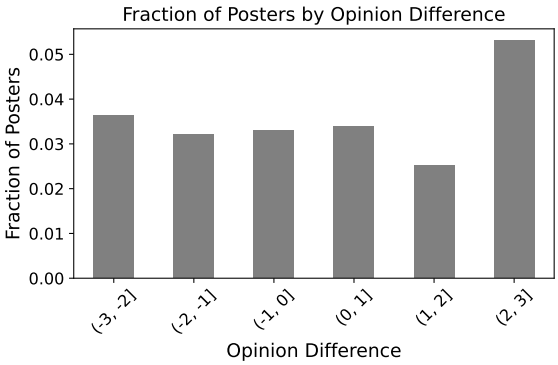
(a) Probability of posting on 4^{th} day



(b) Probability of posting on 5^{th} day



(c) Probability of posting on 6^{th} day



(d) Probability of posting on 7^{th} day

Figure 4.B.3: **Probability of posting**; We look at the probability of a user creating a subsequent post about an asset four, five, six and seven days after their original post in Figures 4.B.3a, 4.B.3b, 4.B.3c, 4.B.3d, respectively. The probability is shown in buckets versus the difference between the author's sentiment in the follow-up post and the average sentiment of peers expressed in the three, four, five, and six days prior to that post, respectively.

We also consider the possibility that there may be another important source of endogeneity driving the choice of certain authors to create a second post rather than others: the fact that the market moves in line with the author's predictions. If this were the case, instead of peer effects, we may be observing author i's reaction to general market developments, rather than true peer effects. To ensure that this is not the case, we include the average return in asset j between author i's two posts $(t-1, t)$. We presents the results in Table 4.B.3 – we observe that our original estimates remain consistent with this control. This implies that peers that post multiple times do not systematically reflect market moves, but indeed present a distinct perspective. The potential choice of post in-line with market moves does

not confound our estimates.

Table 4.B.3: Peer influence in WSB sentiments: Robustness check including average inter-post returns

	Frequent Posters (1)
	Dependent Variable: Investor Sentiment ($\Phi_{i,j,t}$)
Average peer sentiment, $\hat{\Phi}_{-i,j,(t-1)}$ *(predicted)*	0.188 (0.050) ***
$r_{j,t}$	0.990 (0.170) ***
$r_{j,(t-1,t)}$	-0.032 (0.022)
Author & asset controls ($X_{i,j,t}$)	Yes

Notes: this table presents the non-normalized coefficient estimates for the Second Stage of our *Frequent Posters* regression when also including the average return in asset j in the time period between author i's posts, $(t-1,t)$. *** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Temporal robustness check We run several additional robustness checks in order to insure the validity of our identification strategy. One concern may arise from the fact that individuals may be slow to update their opinions, in relation to historical peer posts. In this way, individual i's sentiment about asset j would be affected by peer sentiment before time $(t-1)$ at both times $(t-1)$ and t. In order to tackle this concern we run two different experiments and set a minimum time between $(t-1)$ and t to be at least: i) one day and ii) three days. An additional concern may be whether author i is actually exposed to the sentiments of the peers in question – if the time between $(t-1)$ and t is large, it may be more likely that the author is influenced by outside factors, rather than is constantly exposed to peers during such a prolonged time period. In order to tackle this we run an additional experiment capping the time between posts to be fourteen days. The results for these experiments are presented in Table 4.B.4.

Longer time horizon We use the longer-time period dataset (containing data through mid-2021) to verify whether our findings hold at a longer time horizon.[2] Specifically, we look at

[2]It is worthwhile to note that our data through 2020 is not identical to theirs due to their selection of a different subset of tickers from AlphaVantage and due to different data processing methodologies.

Table 4.B.4: Robustness check for Frequent Posters approach: controlling for the time between author i's posts

	Greater than One Day (1)	Greater than Three Days (2)	Less than Twelve Days (3)
	Dependent Variable: Investor Sentiment ($\Phi_{i,j,t}$)		
Average peer sentiment, $\bar{\Phi}_{-i,j,(t-1,t)}$ *(predicted)*	0.241 *(0.061)* ***	0.195 *(0.071)* ***	0.223 *(0.066)* ***
$r_{j,t}$	1.040 *(0.195)* ***	1.069 *(0.223)* ***	0.930 *(0.190)* ***
Author & asset controls ($X_{i,j,t}$)	Yes	Yes	Yes
Number of obs.	9,712	8,406	5,545
F-statistic	61	53	20

Notes: this table presents the Second Stage OLS estimates for peer influence on WSB limiting the time between user i's posts to be at least one day in Column (1), at least three days in Column (2), and at most fourteen days in Column (3). Variables are not normalized and therefore coefficients are comparable to those in Table 4.B.2. *** Significant at 1% level ** Significant at 5% level * Significant at 10% level

two time periods using the alternative data: i) data through February 1, 2021, and ii) data after February 1, 2021 (the dataset spans through June 27, 2021). GME reached it's peak on January 28, 2021. The results are presented in Table 4.B.5. We note that for the time period ending shortly after the GME short-squeeze, presented in Column (1), the results are consistent with our previous findings. However, in the period following GameStop, the sentiments of peers and recent market moves appear to lose significance, however, users appear more stubborn in their views as visible from the coefficient on $\Phi_{i,j,t-1}$ in Column (2) – user i's own historical sentiment about stock j. The findings underscore the fact that the GME frenzy led to a change in the quality of informational content shared on WSB, as also documented in Bradley et al. (2024). It also appears that users became more stubborn and unwilling to update their views - consistent with the later narrative about the forum as a social movement, rather than a place to discuss meaningful insights about assets.

4.B.4 Evidence of identification strategy – Commenter Network

Distribution of sentiments A potential concern with our approach is whether the sentiments expressed by individuals who are part of the commenters network follow the same

Table 4.B.5: Peer influence: *Frequent Posters* – IV estimates pre- and post-GME

		Dependent Variable: $\Phi_{i,j,t}$	
		Pre-GME (1)	Post-GME (2)
Independent Variables	$\Phi_{i,j,t-1}$	0.228 (0.052) ***	0.297 (0.015) ***
	$\hat{\Phi}_{-i,j,(t-1,t)}$	0.024 (0.008) ***	0.008 (0.012)
	$r_{j,t}$	0.006 (0.003) **	0.002 (0.003)
	$\bar{r}_{j,t}$	0.001 (0.003)	0.005 (0.004)
	$\sigma^2_{j,t}$	0.001 (0.003)	0.009 (0.005) *
	Ticker Fixed Effects	Yes	Yes
	No. Observations:	17,833	9,084
	R^2:	0.13	0.13
	R^2_{adj}:	0.12	0.12

Notes: The dependent variable is individual investor sentiment about an asset, scaled continuously between $(-\infty, \infty)$, is estimated by the individual's previously expressed sentiment about the same asset ($\Phi_{i,j,t-1}$) and a set of market control variables ($r_{j,t}, \bar{r}_{j,t}, \sigma^2_{j,t}$), using OLS. The sentiment of peers ($\hat{\Phi}_{-i,j,(t-1,t)}$) is estimated using data ending immediately after GME hit (spanning through January 31, 2021) in Column (1) and then using data after GME hit its peak in Column (2). The dataset of Semenova et al. (2024) is used for these experiments. Variables are normalized.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

distribution as all submissions on the forum. Figure 4.B.2b presents the distribution of sentiments for those who comment on other's posts, which provides evidence that the sentiment distributions are similar to that of other posters on WSB. This supports the hypothesis that our analysis offers insight into how all individuals on WSB form opinions.

Random network rewiring A concern is whether we effectively control for unobserved ticker characteristics. Similarly to Patacchini & Zenou (2016), we run 'placebo tests', where we replace the composition of an author's peers with a random cohort of people who post on WSB about the same ticker. The random cohort is chosen as follows. We observe how many peers an individual author has. We then select a random sample of the same number of individuals, without replacement, through a random network rewiring (we select posts randomly about the same ticker before the current post). The results are presented in Tables 4.3, column (3). We observe that all the coefficients remain close to their original values, except for the peer effect, which becomes insignificant. This lends credibility to our

peer identification strategy and shows that unobserved factors that influence within ticker variation are not confounding our estimates.

Overidentifying restrictions We cannot directly calculate the J-statistic for our *Commenter Network* approach, since we estimate our IV using observations on several neighbours. We, therefore, take an average of the neighbours past sentiments (transforming the categorical variable into a continuous one) and the average across their neighbour's neighbours sentiments. We use this to compute a J-Statistic with two degrees of freedom.

4.B.5 Followership ties & content consumption on Reddit

Our empirical identification strategy in the *Frequent Posters* approach rests on the premises that users are exposed to random variation in peer sentiments. In this section, we discuss the details of how users are presented with content on Reddit. Upon logging into Reddit, users are presented with a 'home feed'. Historically, the home feed has contained the 'top posts' from the subreddits to which a user has subscribed. Subreddit top posts are not individually tailored to the specific user. Users have several sort options based on whether they prefer to see most recent or most highly rated content.[3]

More recently, Reddit has implemented an algorithm to try and match users to content based on a machine learning algorithm, via the home feed.[4] However, this change has only taken effect recently – during 2021 and after our sample period.[5] Reddit began to experiment with personalization seven years ago, however, focused exclusively on gaming subreddits.[6] However, personalization is focused on identifying *new* subreddits that a user might be interested in, rather than the way that they are exposed to content from subreddits

[3] https://www.reddit.com/r/help/comments/717686/order_of_posts/

[4] https://reddithelp.com/hc/en-us/articles/4402284777364-What-are-home-feed-recommendations-

[5] https://www.reddit.com/r/blog/comments/o5tjcn/evolving_the_best_sort_for_reddits_home_feed/

[6] https://www.reddit.com/r/changelog/comments/5zrtnr/testing_community_recommendations/

they already subscribe to. The explicit goal of personalization is described as an effort to indentify new communities users might be interested in:

> After 15+ years and millions of feedback comments, survey responses, customer interviews, and Mod Council conversations, we know that whether you've been here since the great Digg migration or because you heard about a little community called r/wallstreetbets, we want to help you find communities that you will love on Reddit.

Reddit has, relatively recently, introduced the option to follow individual users, however, following a user means getting exposed to what they post directly to *their own page*, similarly to following an additional subreddit dedicated exclusively to this user. The experience is described as:

> Following is just like subscribing to a subreddit, except the subreddit is your profile page. Reddit recently added the ability to post directly to your profile instead of to a specific subreddit. Posts you post to your profile will be seen on a user's front page feed if they follow you. Outside of that, following does nothing else except your username will be listed in their subscribed subreddits list.[7]

The content viewed on WSB would, therefore, remain consistent for all users regardless of their followership, with random temporal variation, across users. Furthermore, given the fact that the anonymity of the forum is of great appeal, followership ties are rare.[8] We consider how pervasive followership relationships are on Reddit by studying which users actually post to their own profiles (the only way to target content directly at followers). We look through all 42,036 authors who create posts about individual tickers within our sample

[7] reddit.com/r/NoStupidQuestions/comments/9dzp9y/what_does_following_someone_on_reddit_do/

[8] https://www.reddit.com/r/NoStupidQuestions/comments/7lkwqs/do_reddit_users_actually_follow_other_people/

and observe that less than 3% of users create content on their own individual user profile pages prior to our data cutoff time, demonstrating the relative lack of content generated for followers and the insignificance of followership relationships on Reddit. Furthermore, we observe that content posted to WSB user's own profile (which we retrieve) is generally unrelated to investment – investment advice is typically shared on investment-related forums to reach a targeted audience. We test the sensitivity of our results to the users that post to their own profiles remaining in our sample by removing them and rerunning the *Frequent Posters* estimation procedure: the results remain unchanged when these users are removed.

We conclude that individual users were exposed to WSB content based on the content on the forum that was most recent and popular at the time of their logging on, rather than based on personal preference. This, in turn, allows us to assert that users are exposed to random, temporal variation in peer sentiment.

Chapter 5: Social Dynamics and Asset Prices

Chapter 4 documents two important factors that impact investor sentiment – peer effects and extrapolation. This chapter explores several models to help understand the patterns of returns when these mechanisms of sentiment formation among investors are present. Section 5.1 considers the rational investor deriving their expectations about future asset returns based on extrapolation and peer effects; liquidity is provided by noise traders – the model allows us to consider temporary deviations from market efficiency driven by behavioural patterns. Section 5.2 considers a modified model for extrapolative bubble dynamics, where we modify the framework to incorporate a social component. Finally, Section 5.3 considers a complex systems framework to analyze regime-shifts and tipping points in market dynamics. The focus of each section within this chapter is to understand how different estimates for α and β (explored in the previous chapter) alter market dynamics.

5.1 Equilibrium Model with Social Shocks

Our proposed model is motivated by a growing literature on strategic information complementarities (Hellwig & Veldkamp 2009, Zenou 2016), as well as studies of diagnostic expectations (Bordalo et al. 2021). Investors trade on the momentum of the stock price, against a supply of shares provided by noise traders. Our inclusion of a social component subsequently induces persistence in asset demand over time, which leads to reversal in future returns.

General setup We analyse the price of one asset traded by N investors, indexed by i. Each investor derives Constant Absolute Risk Averse (CARA) utility from consuming c, $U_i(c_i) = -\exp(-\gamma c_i)$, where γ is the constant absolute rate of risk aversion – the model setup

is consistent with various behavioural models, models for bubble formation, and is justified by empirical observations on the relationship between investor sentiment and volatility (Bordalo et al. 2021, Barberis et al. 2018). We do not include any discounting in investor decision-making, but assume they evaluate an asset according to a log-normal distributed value $v = \log(V)$ with expectation $\mathbb{E}_i(v)$ and variance $\sigma_i^2(v)$. In the static model, investor i purchases ϕ_i shares at the current market log-price p to optimise the mean-variance objective function

$$\mathcal{L}(\phi_i) = [\mathbb{E}_i(v) - p]\phi_i - \frac{\gamma}{2}\sigma_i^2(v)\phi_i^2, \tag{5.1}$$

$$\Rightarrow \phi_i^* = \frac{\mathbb{E}_i(v) - p}{\gamma\sigma_i^2(v)}, \tag{5.2}$$

where an asterisk denotes the value that maximises objective $\mathcal{L}$. In this way, we distinguish between beliefs about value $\mathbb{E}_i(v)$ from investor i's decision to buy amount ϕ_i. Eq. 5.2 yields a familiar expression for asset demand in equilibrium: namely a ratio of expected net returns over their variance.

Assuming that asset supply originates from noise traders, S, as in Bordalo et al. (2021), common variance $\sigma^2 = \sigma_i^2(v)$ and averaging expected values $\mathbb{E}(v) = 1/N \sum \mathbb{E}_i(v)$, we can re-arrange Eq. 5.2 to yield the following expression for the market-clearing price p:

$$p = \mathbb{E}(v) - \frac{S}{N}\gamma\sigma^2. \tag{5.3}$$

Eq. 5.3 accounts for the price level by investor's average expected value of the asset, in addition to their ability to absorb the exogenous level of assets supplied. This ability depends on the depth of the investor pool – reflected by the number of investors N – as well as their risk appetite $\gamma\sigma^2$. In this simple market, the price increases with expected value, and decreases with supply. Appendix 5.A further links this model framework to complementarities in

asset demand.

Granular shocks An investor's demand may have some idiosyncratic preference which is not captured by common factors. For example, an investor may place particular confidence in products he enjoys using, or admire the corporate strategy of certain company leaders. Such sentiments would manifest as idiosyncratic, heterogeneous investor demand, where e_i is the idiosyncratic component of i' asset demand. Under the assumption that these shocks have a finite variance and mean zero, they should average out to zero by the Central Limit Theorem. However, consider a scenario where some investors have different levels of importance s_i for aggregate demand:

$$S = \sum_{i=1}^{N} s_i \phi_i^*, \tag{5.4}$$

$$\Rightarrow p = \sum_{i=1}^{N} s_i \mathbb{E}_i(v) - \gamma \sigma^2 S + \sum_{i=1}^{N} s_i e_i, \tag{5.5}$$

where weights $\sum_i s_i = 1$ for demand to equal supply. This could be the case for several reasons: investors could have different amount of capital or, of greater interest to this chapter, some investors may have *more sway* in forming public opinion than others. The importance of key players in a social context has been explored in several economic settings – see Zenou (2016) for a thorough overview. Typically, the most central nodes in a social network have the ability to quickly diffuse information and, therefore, have a high influence on others.

We justify this weighting scheme by the fact that certain users on WSB have a disproportionate effect in shaping the broader discourse. Indeed, WSB is structured to promote viral content, and we would expected a consensus to be formed by key players – or, rather, around key submissions. If the distribution of importance does not have a finite variance – i.e. it is 'heavy-tailed' – then the idiosyncratic shocks would not average out to zero. We use this framework to evaluate the impact of viral content from WSB on market returns in

Chapter 6.

5.1.1 Equilibrium price dynamics with peer effects

To study the joint dynamics of an asset's price and demand by social investors, we treat aggregate asset demand $\phi = \sum_i \phi_i$ and log price p as state variables for a dynamic system, indexed by time t. In doing so, we assume that cumulative demand ϕ_t reflects a difference between individual valuations of the asset and the price. We distinguish between two independent components of individual valuations: the private signal of individuals $g(b_{i,t})$ and the signal individuals draw from observations of peers $f(\phi_{i,t})$. Aggregate asset demand and price are

$$\phi_t = \frac{\mathbb{E}_t[g(b_{i,t})] + \mathbb{E}_t[f(\phi_{i,t})] - p_t}{\gamma\sigma^2}, \tag{5.6}$$

$$p_t = \mathbb{E}_t[g(b_{i,t})] + \mathbb{E}_t[f(\phi_{i,t})] - \frac{S_t}{N}\gamma\sigma^2. \tag{5.7}$$

Cumulative demand is, therefore, the difference between individual valuations and the market-clearing price, normalized by risk-aversion. The market-clearing price, on the other hand, is the difference between individual valuations and the rate of asset supply, normalized by the number of investors N and their risk appetite $\gamma\sigma^2$, similar to Eq. 5.3.

A focus of this chapter is the relationship between valuations $g(b_{i,t})$ and the social component $f(\phi_{i,t})$. Studies in behavioural finance suggest different expectation formation mechanisms that ultimately deviate from rational expectations (Barberis et al. 2018, Bordalo et al. 2021). We combine two such features, which we explore and document in the previous chapter, in Assumptions 1 and 2 to formalize the framework for our modeling exercise.

Persistent demand The mechanism by which past demand enters current asset demand is by the complementarity in investor payoffs – which empirically manifest as peer effects

(documented in Chapter 4). Investor i's payoff to holding the asset is assumed to increase linearly in average asset demand by others.

Assumption 1 (Persistent Demand)**.** *Social investor i's expectation of future returns is linearly increasing in average asset demand by others:* $f(\phi_{i,t}) = \alpha\phi_{t-1}$*, where* $\phi_{t-1} = 1/N \sum_i \phi_{i,t-1}$ *is average asset demand.*

Mechanical Extrapolation We assume that investors partially trade on the momentum of the asset's price, which we study in Chapter 4 as extrapolation. The functional form of $\mathbb{E}_t[g(b_i)]$ is specified in Assumption 2.

Assumption 2 (Mechanical Extrapolation)**.** *The average investor projects past price increases into the future using the updating rule:*

$$\mathbb{E}_t[g(b_{i,t})] = p_t + \beta(p_t - p_{t-1}), \tag{5.8}$$

where β *captures a fixed degree of price extrapolation.*

System for price and demand Combining Assumptions 1 and 2 into Eqs. 5.6-5.7 yields demands and returns $r_t = p_t - p_{t-1}$:

$$\phi_t = \frac{\alpha\phi_{t-1} + \beta r_t}{\gamma\sigma^2}, \tag{5.9}$$

$$r_t = -\frac{\alpha}{\beta}\phi_{t-1} + \frac{S_t\gamma\sigma^2}{\beta N}. \tag{5.10}$$

In this scenario, asset demand and returns are determined simultaneously. The first mechanism is through market clearing, where demand has to adjust to supply. The second is the

adjustment of the expected value for the asset to the realised return, through β, and the social signal, through α.

As a result, returns are accounted for by current and past asset demand:

$$r_t = \frac{\gamma\sigma^2}{\beta}\phi_t - \frac{\alpha}{\beta}\phi_{t-1}, \tag{5.11}$$

since supply must equal demand at time t. This equation uncovers several important mechanisms at play. Returns are related positively to current demand ϕ_t through market clearing – supply S is exogenous and must meet current demand. Higher demand drives up returns. Returns are, however, negatively related to past demand ϕ_{t-1} through the expected value of an asset – a *valuation mechanism*. If there is a positive social signal, investors still value an asset highly, even in the presence of low returns.

To explain the basic intuition, we consider the following scenarios: i) one where the asset has a positive return r_t and no social signal ϕ_{t-1}, and ii) one where investors observe a positive social signal ϕ_{t-1}. In scenario (i), demand is driven by the extrapolation component alone – investors believe that returns will continue to increase based on the current trend. In scenario (ii) on the other hand, investors do not require a large return to demand the asset – positive past sentiment drives current demand ϕ_t. Under exogenous supply, a strong positive signal from peers means that the extrapolated return is *less* important in justifying a higher price. The underlying reasoning relies on the fact that the system is in equilibrium. Therefore, both returns and sentiments have adjusted to reflect a new steady-state, where sentiments are at a certain level ϕ_t.

Finally, we observe that the ratios of the coefficients, γ/β and α/β, play an important role. β effectively anchors the demand of investors in reality – a greater value of β implies that social signals carry less weight, and investors focus on price trends to forecast and expect asset values to grow at some constant rate. As β decreases, returns are determined

more by social forces – hype from peers, rather than past performance, now justifies returns and demand. In our data, we observe that α is roughly two times β – individuals weight the sentiments of peers *more heavily* than recent returns. σ^2 serves to taper the impact of sentiment, since investors are less certain in their signal and demand less of the asset.

5.1.2 Persistent fluctuations

The reversal in returns is an important feature that emerges from social contagion in investors' price expectations. If large enough, these can produce bubbles in asset prices: initial momentum from positive news creates a price run-up, before an absence of news creates a drought of new asset demand. The subsequent price crash carries on its own momentum. We can treat demand as a latent variable to see these oscillations manifest in return data. Substituting lagged demand into the equation for returns, and iterating infinitely yields

$$r_t = -\sum_{T=1}^{t} \left(\frac{\alpha}{\gamma \sigma^2} \right)^T r_{t-T} + \frac{S_t \gamma \sigma^2}{\beta N} \tag{5.12}$$

as long as $\alpha/\gamma\sigma^2 < 1$, so that the contribution of demand fluctuations to returns converges to zero over time. This is an autoregressive model with infinite lags, where the coefficients decrease exponentially with lag size T. Without any knowledge of asset demand, the second term encapsulates an unobservable error term, which the model links to exogenous changes in the asset's supply. Eq. 5.12 demonstrates that an exogenous increase in returns at time t is followed by a smaller decrease in $t+1$. This oscillation persists indefinitely, and would converge to zero rapidly if the social signal $\alpha/\gamma\sigma^2$ is sufficiently small.

5.2 Bubbles with peer effects

In addition to the setting explored above, we demonstrate how peer effects are relevant in modeling bubble dynamics through incorporating them in an extension of Barberis et al.

(2018). We highlight the relevant model details below, however, direct the reader to the original paper for the full model setup. In the original model, extrapolators determine their demand from a 'fundamental signal' with weight w_i, as well as an 'extrapolation signal' with weight $(1-w_i)$, and trade with fundamental traders in the market. The demand function for extrapolators with non-varying temporal weights is:

$$w_i \frac{F_t}{\gamma\sigma^2} + (1-w_i)\frac{M_t}{\gamma\sigma^2}. \tag{5.13}$$

Barberis et al. (2018) define M_t (X_t in the original text) as:

$$\begin{aligned} M_t &= (1-\theta_E)\sum_{k}^{t-1}\theta^{k-1}(P_{t-k}-P_{t-k-1}) + \theta^{t-1}X_1, \\ &= (1-\theta_E)(P_{t-1}-P_{t-2}) + \theta_E M_{t-1}. \end{aligned}$$

where θ_E is the weight placed on recent versus older price changes, and is between zero and one.

We propose to modify extrapolator signal to incorporate a social component, M_t^s:

$$M_t^s = (1-\theta_E)\phi_t + \theta_E M_{t-1}^s, \tag{5.14}$$

where ϕ_t is the average sentiment determined from past price returns (extrapolation), as well as a past expressed sentiments (persistent demand driven by peer effects),

$$\phi_t = \beta^*(P_{t-1}-P_{t-2}) + \alpha^*\phi_{t-1}, \tag{5.15}$$

where β^* and α^* sum to one.

We use our estimates for β and α to compare resulting bubble dynamics in the presence

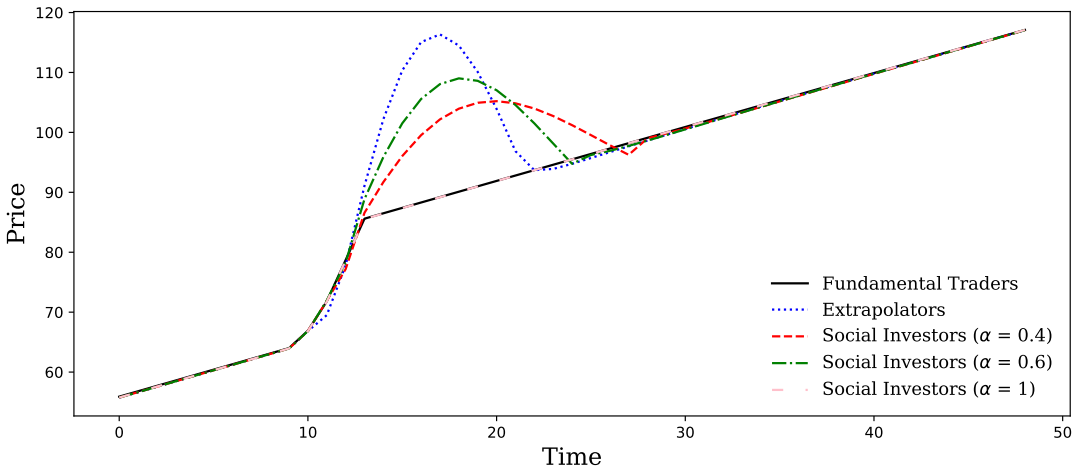
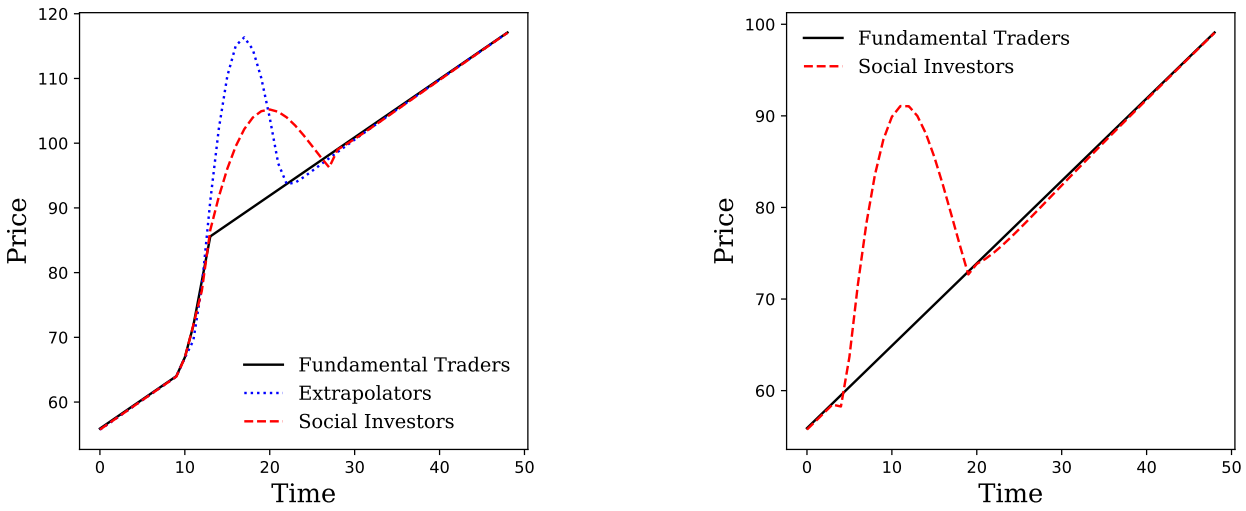


Figure 5.1: **Bubbles with Different Parameter Values for α^*, β^***; We choose initial parameters similar to those from Figure (1) in Barberis et al. (2018): $w_i = 0.1$, $\sigma^2 = 3$, $\gamma = 0.1$, $\theta_E = 0.1$; extrapolators / social investors make up 70% of investors, the remainder are fundamental traders; the quantity of the asset is set to 1. In Figure 5.2a, we choose that the fundamental value of the asset remains unchanged except for periods 11-14 when information is revealed resulting in an increase in the future dividend of the asset by 2,4,6,6 (respectively).

of peer effects to the original findings in Barberis et al. (2018) – our estimates demonstrate that social investors on WSB place a relative weight α^* of 0.6 on the sentiments of peers and a relative weight β^* of 0.4 on recent returns. Figure 5.1 demonstrates that the values for α^*, β^* control how long of a memory investors have. As α^* increases, we observe that the bubble takes a longer period of time to form and dissipate.

We run several additional simulations of the modified model for bubbles and present the results in Figure 5.2. The simulations demonstrate two things. As documented previously, in an extrapolative setting, the existence of a social signal makes the bubble formation process have longer memory; the bubble takes longer to form, and has a less defined peak. Second, the modification allows a mechanism for bubbles to form as a result of a purely social signals, thereby introducing the potential for 'animal spirits' among investors that result in bubble-like dynamics.



(a) Bubbles with Extrapolation, Sentiments

(b) Sentiment-Driven Bubbles

Figure 5.2: **Bubbles with Social Investors**; We simulate our modified version of the model for bubbles with extrapolation (Barberis et al. 2018). We choose initial parameters similar to those from Figure (1) in Barberis et al. (2018): $w_i = 0.1$, $\sigma^2 = 3$, $\gamma = 0.1$, $\theta_E = 0.1$; extrapolators or social investors make up 70% of investors, the remainder are fundamental traders; the quantity of the asset is set to 1. In Figure 5.2a, the fundamental value of the asset remains unchanged except for periods 11-14 when information is revealed resulting in an increase in the future dividend of the asset by 2,4,6,6 (respectively). In Figure 5.2b, we impose a comparable sentiment shocks of 6 in periods 5,6,7.

5.3 Complex systems model with social contagion

The previous two sections present quantitative finance models for how extrapolation and peer effects can impact asset prices. In this section, we take a complex systems perspective. Despite there being a long-standing history of studying social movements with respect to their tipping points (Lamberson et al. 2012, Moser & Dilling 2007), the connection to financial markets is under-explored. This section's contribution rests on bringing relevant complex systems techniques to shed light on how social media can impact the financial markets.

Consider a market for one asset with two types of participants: 'hype' investors, who buy Y shares, and 'non-hype' investors, who buy quantity S. Their total demand is equal to the number of shares outstanding,

$$Q = Y + S.$$

We propose a model in discrete time t where hype investors observe the behaviour of others, as well as the asset's returns, before updating their demand for the asset to maximise their expected utility. The goal is to study the stability of a market with social dynamics, given that hype investors are able to move price. The key endogenous variable is asset demand, which governs the propensity for hype investors to buy or sell the asset, thus shifting demand Y_t.

Total hype investor demand Y_t is composed of i) the average demand across all active hype investors, $d_t \in [-1, 1]$, ii) the average purchasing power of an individual hype investor M/p_t, where M are funds available to an average hype investor, and iii) the total number of hype investors N:

$$Y_t = \frac{M}{p_t} N d_t. \tag{5.16}$$

We largely treat total funds available to hype investors MN as exogenous, and focus on their decision to buy or sell the asset, reflected by their average demand d_t. We endogenize demand d_t to change with the asset's price, as well as the degree of coordination among hype investors.

Individual investor demand We formalise a functional form for a given hype investor's demand $d_{i,t+1}$, indexed by i. The utility an investor derives from buying, or selling, the asset is a function of other hype investors' average expressed *sentiment* ϕ_t, asset return $r_t \equiv p_t/p_{t-1} - 1$, as well as an error ϵ_i:

$$U_i(d_{i,t+1} = +1) = \epsilon_i^+ + \alpha\phi_t + \beta r_t - \gamma r_t^2, \quad U_i(d_{i,t+1} = -1) = \epsilon_i^- - \alpha\phi_t - \beta r_t - \gamma r_t^2. \tag{5.17}$$

where β is a trend following (extrapolation) parameter, α is the network spillover effect from peers (peer effects), and γ captures aversion to large swings in price. In this manner, the utility of buying(selling) the asset increases(decreases) with peer *sentiment*, as the investor seeks to align their position with that of other hype investors. They must rely on sentiment, because we assume that hype investors are unable to communicate the size of their position, and instead only share their decision to buy or sell the asset. Utility also increases(decreases) with recent price returns, and decreases(decreases) with return volatility. This imparts a trend-following tendency for hype investor's decisions, but also an aversion to large swings in price. Residuals ϵ_i^+ and ϵ_i^- capture unobserved variation in utility among various hype investors when buying or selling the asset.

While we do not observe demand on WSB, we observe sentiments. We interpret sentiments as a binary random variable with which a user expresses their preference to buy ($\phi_{i,t+1} = 1$) or sell ($\phi_{i,t+1} = -1$) the asset. This allows us to determine the probability for a hype investor to express a given sentiment, based on observed peer sentiment and asset returns.

ϵ_i^+ and ϵ_i^- refer to a random, exogenous element in individual preferences. We make Assumption 3 about ϵ_i^+ and ϵ_i^-, which will allow us to find an expression for aggregate sentiment.

Assumption 3. *ϵ_i^+ is independent and identically distributed, following a type-I Extreme Value (EV) distribution. The same holds for ϵ_i^-.*

Assumption 3 is standard in the literature (Bouchaud 2013), and justifiable in our present context since investors make a choice governed by a maximisation process.

An assumption on the distribution of the idiosyncratic terms ϵ_i^+ and ϵ_i^- (Assumption 3) allows us to derive an expression for the probability of bullish versus bearish sentiment. We

write the probability of bullish sentiment as

$$\begin{aligned}
\mathrm{P}[U_i(+1) > U_i(-1)] &= \mathrm{P}(\epsilon_i^+ + \alpha\phi_t + \beta r_t - \gamma r_t^2 > \epsilon_i^- - \alpha\phi_t - \beta r_t - \gamma r_t^2), && (5.18)\\
&= \mathrm{P}(\epsilon_i^- < \epsilon_i^+ + 2\alpha\phi_t + 2\beta r_t), && (5.19)\\
&= \int_{-\infty}^{+\infty} \mathrm{P}(\epsilon_i^- < \epsilon_i^+ + 2\alpha\phi_t + 2\beta r_t | \epsilon_i^+) \times p(\epsilon_i^+) d\epsilon_i^+, && (5.20)\\
&= \int_{-\infty}^{+\infty} \exp\left[-\exp\left(-\frac{\epsilon_i^+ + 2\alpha\phi_t + 2\beta r_t}{2\lambda}\right)\right] \times \frac{1}{2\lambda} \exp\left[-\frac{\epsilon_i^+}{2\lambda} - \exp\left(-\frac{\epsilon_i^+}{2\lambda}\right)\right] d\epsilon_i^+, && (5.21)\\
&= \frac{\exp\left(\frac{\alpha\phi_t + \beta r_t}{\lambda}\right)}{\exp\left(\frac{\alpha\phi_t + \beta r_t}{\lambda}\right) + 1}, && (5.22)
\end{aligned}$$

where step 5.21 follows from the fact that we are integrating the cumulative density of ϵ_i^-, with scale 2λ, over the support of ϵ_i^+, with marginal density with marginal density $p(\epsilon) = \exp[-\epsilon/2\lambda - \exp(-\epsilon/2\lambda)]/2\lambda$. The final simplification comes from the fact that $U(d_{i,t+1} = +1)$ and $U(d_{i,t+1} = -1)$ are symmetric in parameters α, β and γ.

This exercise yields the following:

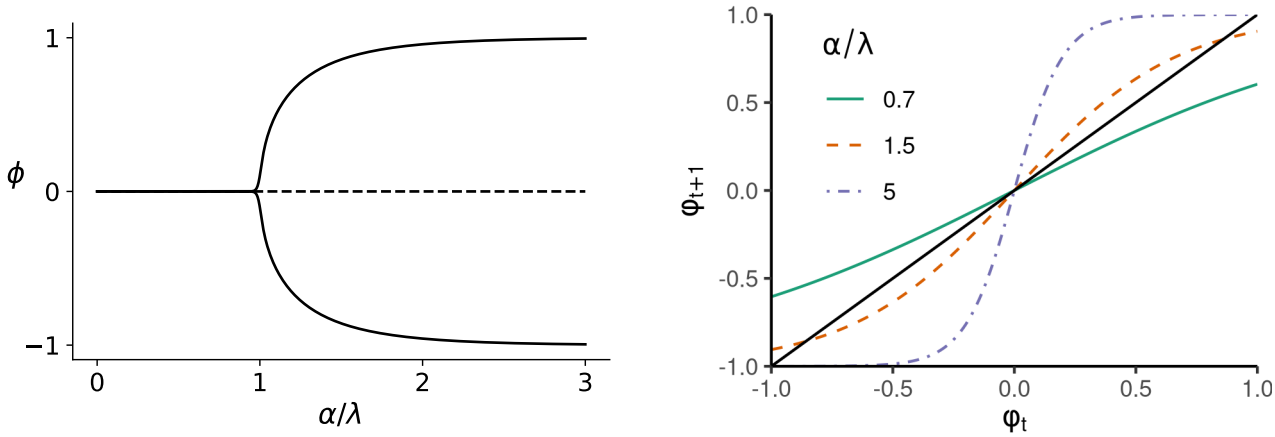
$$\phi_{i,t+1} = \begin{cases} +1, & \text{with probability } \mathrm{P}(\phi_{i,t+1} = +1) = \frac{\exp[(\alpha\phi_t + \beta r_t)/\lambda]}{\exp[(\alpha\phi_t + \beta r_t)/\lambda] + 1}, \\ -1, & \text{with probability } \mathrm{P}(\phi_{i,t+1} = -1) = \frac{1}{\exp[(\alpha\phi_t + \beta r_t)/\lambda] + 1}. \end{cases} \tag{5.23}$$

Investor i's sentiment $\phi_{i,t+1}$ is what we observe on social media, and is assumed to be an honest evaluation of their investment decision $d_{i,t+1}$. Parameter λ is a scale parameter for the breadth of the error distributions, assumed to follow a Gumbel distribution, and specifies the amount heterogeneity in individual investment preferences (Bouchaud 2013).

Aggregate hype investor buying intensity We aggregate sentiments into an average measure ϕ_{t+1} using Eq. 5.23:

$$\begin{aligned}\phi_{t+1} &= \frac{1}{N}\sum_{i=1}^{N} \mathrm{P}(\phi_{i,t+1} = +1) - \mathrm{P}(\phi_{i,t+1} = -1) \\ &= \frac{\exp\left[(\alpha\phi_t + \beta r_t)/\lambda\right] - 1}{\exp\left[(\alpha\phi_t + \beta r_t)/\lambda\right] + 1} \\ &= \tanh\left[\frac{\beta r_t + \alpha\phi_t}{\lambda}\right]. \end{aligned} \tag{5.24}$$

Thus, Eq. 5.23 determines the evolution of average sentiments as a function of past sentiments, in addition to past returns.



(a) Steady states of the hyperbolic tangent function; we plot ϕ when the system has converged after 100 time steps for different values of α/λ and ϕ_0.

(b) Hyperbolic tangent function in ϕ_t; we plot $\phi_{t+1} = \tanh(\alpha\phi_t/\lambda)$ for three values of α/λ. The function crossing the 45 degree line (solid black) denotes a steady state.

Figure 5.3: **Properties of the hyperbolic tangent;** Figure 5.3a shows the bifurcation of the hyperbolic tangent at $\alpha/\lambda = 1$. When $\alpha/\lambda < 1$, $\phi = 0$ is the unique, stable steady state. When $\alpha/\lambda > 1$, two new stable steady states emerge, ϕ^+ and ϕ^-, and $\phi = 0$ becomes unstable. Figure 5.3b, shows the temporal behaviour of the hyperbolic tangent function for different α/λ. When $\alpha/\lambda = 0.7$, the function only crosses the 45 degree line (solid black) once at the origin. For values of 1.5, in dashed orange, and five, in dot-dashed purple, it crosses the 45 degree line in three places, indicating the emergence of two new steady states.

Properties of the hyperbolic tangent function The function in Eq. 5.24 is known as the *hyperbolic tangent* function, which has received significant attention in the area of strategic decision-making (Brock & Durlauf 2001, Bouchaud 2013). We highlight the key properties of the hyperbolic tangent function in our application, as they are important components of

our model's stability. When α/λ is between zero and one, the function produces one stable steady state for $\phi_{t+1} = \phi_t$ at zero. Two stable steady states, and one unstable one at zero, emerge when α/λ is greater than one, as shown in Figure 5.3a. Figure 5.3 presents a visualization of the pitchfork bifurcation, as well as the temporal evolution, of the hyperbolic tangent function (Hommes 2013). The stability of the system is critically dependent on α/λ. When the social signal $\alpha\phi_t$ is high relative to noise λ, investor coordination is effective and all hype investors converge on a similar investment strategy. On the other hand, if α/λ is low, then noise overwhelms the social signal – coordination fails and the zero steady state is the only one that is ever reached.

Hype investor demand and market moves We complete our stylised model by linking hype investor sentiments to price returns r_{t+1}. Recall that, under market clearing, the shares held by hype and non-hype investors must equal outstanding shares Q. For our purposes, we make a simplifying assumption that non-hype investors keep the nominal value invested in the asset fixed, so that

$$\Delta p_t S_t = 0.$$

This enables us to study the emergent price dynamics due to social dynamics in isolation (even though our findings hold when this assumption is relaxed). Under this assumption, the dynamics for price returns are:

$$r_{t+1} = \underbrace{\frac{MN}{p_t Q}}_{\text{(1) Capacity}} \times \underbrace{\Delta\phi_{t+1}}_{\text{(2) Sentiment}} . \tag{5.25}$$

Eq. 5.25 follows from the first difference of the market capitalization:

$$\Delta(p_{t+1}Q) = \Delta(p_{t+1}Y_{t+1}) + \Delta(p_{t+1}S_{t+1}) = MN\Delta\phi_{t+1}, \tag{5.26}$$

where equality holds because $\Delta(p_{t+1}S_{t+1}) = 0$ and the expression for $p_{t+1}Y_{t+1}$ in Eq. 5.16. Rearranging, and substituting $r_{t+1} = \Delta p_{t+1}/p_t$ yields the desired result.

Component (1) in Eq. 5.25, labelled 'Capacity', is a market depth term, measuring how much market power hype investors hold, as a share of the total market capitalization of the asset. Term (2) captures the change in their aggregate sentiment, which reflects underlying trades of hype investors as they respond to lagged returns and sentiments.

Linking asset returns to sentiments is the main feature of the model, for which we make an assumption for the process that generates sentiments as a function of individual hype investor asset demands. This is relevant to our use of social media data, because demands are unobserved.

5.3.1 Market stability in the presence of social dynamics

In the absence of any fundamental news, our model with hype investors generates the following dynamic system:

$$\vec{\phi}_{t+1} = \tanh\left[\frac{\beta r_t + \alpha\phi_t}{\lambda}\right], \tag{5.27}$$

$$\vec{r}_{t+1} = C(\phi_{t+1} - \phi_t), \tag{5.28}$$

where we simplify our initial expression for returns by defining the relative capacity of hype investors $C = MN/p_tQ$. The key dynamic of interest is the strategic coordination observed among hype investors at a given point in time, ϕ_t, and its impact on asset returns. Therefore, we focus on parameters α, β, and C.

Eqs. 5.27 and 5.28 are two simultaneous difference equations, with sentiment ϕ_t and returns r_t as state variables. Importantly, the evolution of aggregate sentiment ϕ_{t+1} is non-linear. The system produces either one or three steady states, depending on parameter values. In any steady state, returns $r_{t+1} = r_t$ must be zero. The steady state values for

sentiments ϕ are governed by α/λ, illustrated by the inset in Figure 5.5a: one at zero, then two more at $\{\phi^+, \phi^-\}$ when $\alpha/\lambda > 1$, which do not have a closed form solution. This yields three steady states for the system as whole when $\alpha/\lambda > 1$, which we denote as (sentiment, return) tuples: i) $(0, 0)$, ii) $(\phi^+, 0)$, and iii) $(\phi^-, 0)$.

Determining the steady states in a dynamic system is critical for understanding how the system evolves over time. As the name implies, when a dynamic system is in its steady state, it will remain there: in our case, sentiments and asset returns will note change. However, it is not obvious how the system behaves outside its steady state(s). For example, if the asset's return is non-zero due to an external shock, or if a rumour rips through the hype investor population and exogenously changes sentiments. To tackle this, we examine the stability of each steady state individually.

We proceed by finding the Jacobean matrix for the steady states listed above: i) $(0, 0)$, ii) $(\phi^+, 0)$, and iii) $(\phi^-, 0)$. We then examine the stability of each steady state using the eigenvalues of the Jacobian evaluated at the steady states. We show that the key parameters governing the system are i) the consensus parameter α, the degree to which investors' sentiments mimic those of their peers, ii) the capacity parameter C, iii) the degree of trend following β, and iv) the noise parameter λ.

Finding the steady states We outline the steps taken to find our system steady states. By definition, r_t must be zero at the steady state. This means that steady states of the dynamic system in Eqs. 5.27-5.28 are those for which ϕ_t is a solution to Eq. 5.27, which is the hyperbolic tangent function displayed in the upper right hand corner of 5.5a. Zero is a unique steady state when $\alpha/\lambda < 1$, and two further steady states emerge when $\alpha/\lambda > 1$. Those two additional steady states are solved for numerically in all simulation exercises (using a solving algorithm).

In a discrete time system, the type of stability around a steady state is dependent on

the eigenvalues of the Jacobian matrix at the steady state. When a system has multiple eigenvalues, different behaviours can emerge.

Eqs. 5.27-5.28 can be rewritten as:

$$\vec{\phi}_{t+1} = \tanh\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right],$$
$$\vec{r}_{t+1} = C(\tanh\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right] - \phi_t),$$

We calculate the Jacobian through taking the derivatives of the above system as follows:

$$J = \begin{bmatrix} \frac{\partial \phi_{t+1}}{\partial \phi_t} & \frac{\partial \phi_{t+1}}{\partial r_t} \\ \frac{\partial r_{t+1}}{\partial \phi_t} & \frac{\partial r_{t+1}}{\partial r_t} \end{bmatrix},$$

which is equivalent to:

$$J = \begin{bmatrix} \alpha/\lambda \cdot \operatorname{sech}^2\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right] & \beta/\lambda \cdot \operatorname{sech}^2\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right] \\ C\left(\alpha/\lambda \cdot \operatorname{sech}^2\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right] - 1\right) & C\beta/\lambda \cdot \operatorname{sech}^2\left[\frac{\beta r_t + \alpha \phi_t}{\lambda}\right] \end{bmatrix},$$

and subsequently computing the values at the steady states $(0, 0)$ and $(\phi, 0)$.

The Jacobian matrices at the different steady states are

$$J(0,0) = \begin{bmatrix} \alpha/\lambda & \beta/\lambda \\ C(\alpha/\lambda - 1) & C\beta/\lambda \end{bmatrix}, \quad J(\phi,0) = \begin{bmatrix} \alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) & \beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) \\ C\left(\alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - 1\right) & C\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) \end{bmatrix},$$

where $\phi \in \{\phi^+, \phi^-\}$.

The eigenvalues for the steady state at $J(0,0)$ is subsequently determined by:

$$\det\begin{bmatrix} \alpha/\lambda - x & \beta/\lambda \\ C(\alpha/\lambda - 1) & C\beta/\lambda - x \end{bmatrix} = 0,$$

$$(\alpha/\lambda - x)(C\beta/\lambda - x) - C\beta/\lambda \cdot (\alpha/\lambda - 1) = 0,$$

$$x^2 - (\alpha/\lambda + C\beta/\lambda)x + C\beta/\lambda \cdot \alpha/\lambda - C\beta/\lambda \cdot (\alpha/\lambda - 1)) = 0,$$

$$x^2 - (\alpha/\lambda + C\beta/\lambda)x + C\beta/\lambda = 0,$$

where we use non-standard notation for eigenvalue of x.

The corresponding eigenvalues for steady state $(0,0)$ are

$$x_1 = \frac{C\beta/\lambda + \alpha/\lambda + \sqrt{(C\beta/\lambda + \alpha/\lambda)^2 - 4C\beta/\lambda}}{2}, \tag{5.29}$$

$$x_2 = \frac{C\beta/\lambda + \alpha/\lambda - \sqrt{(C\beta/\lambda + \alpha/\lambda)^2 - 4C\beta/\lambda}}{2}. \tag{5.30}$$

The eigenvalues for the steady state at $J(\phi,0)$ is subsequently determined by:

$$\det\begin{bmatrix} \alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - x & \beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) \\ C\left(\alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - 1\right) & C\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - x \end{bmatrix} = 0,$$

$$x^2 + (\alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - x)(C\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - x)$$

$$-(\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda))(C\left(\alpha/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda) - 1\right)) = 0,$$

$$x^2 + C\alpha\beta/\lambda^2 \cdot \operatorname{sech}^4(\alpha\phi/\lambda) - x((C\beta + \alpha)/\lambda) \cdot \operatorname{sech}^2(\alpha\phi/\lambda)$$

$$-C\alpha\beta/\lambda^2 \cdot \operatorname{sech}^4(\alpha\phi/\lambda) + (C\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda)) = 0,$$

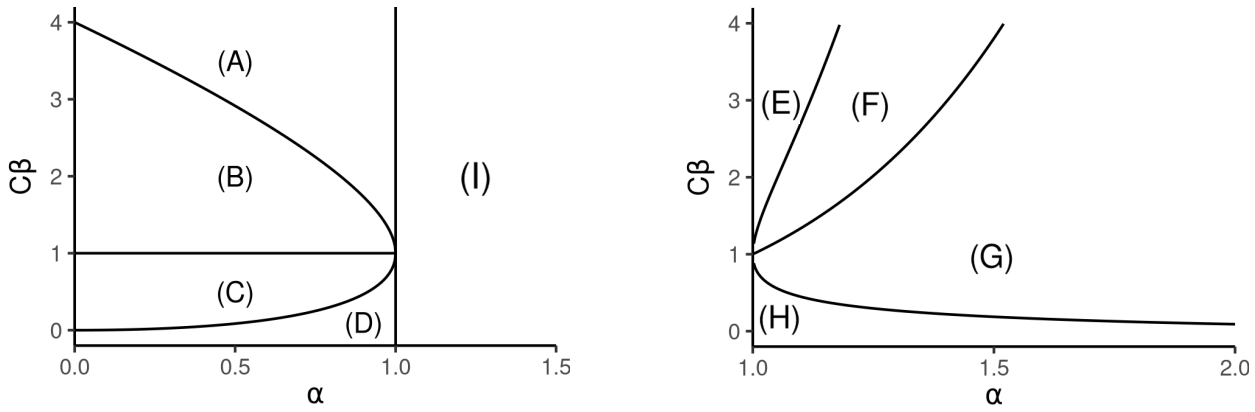
$$x^2 - x((C\beta + \alpha)/\lambda) \cdot \operatorname{sech}^2(\alpha\phi/\lambda) + (C\beta/\lambda \cdot \operatorname{sech}^2(\alpha\phi/\lambda)) = 0,$$

where we use non-standard notation for eigenvalue of x. The eigenvalues at $(\phi, 0)$ are

$$x_1 = \frac{\text{sech}^2(\alpha\phi/\lambda)(C\beta+\alpha)/\lambda + \sqrt{\text{sech}^4(\alpha\phi/\lambda)((C\beta+\alpha)/\lambda)^2 - 4C\beta/\lambda\text{sech}^2(\alpha\phi/\lambda)}}{2}, \quad (5.31)$$

$$x_2 = \frac{\text{sech}^2(\alpha\phi/\lambda)(C\beta+\alpha)/\lambda - \sqrt{\text{sech}^4(\alpha\phi/\lambda)((C\beta+\alpha)/\lambda)^2 - 4C\beta/\lambda\text{sech}^2(\alpha\phi/\lambda)}}{2}. \quad (5.32)$$

These expressions trace out different regions of stability as a function of $C\beta$ and α in Figure 5.4. We drop λ in the next stage of our analysis, as it can be treated as a constant that affects parameters α and $C\beta$ equally – therefore exploring variations in λ is subsumed into varying α and $C\beta$.

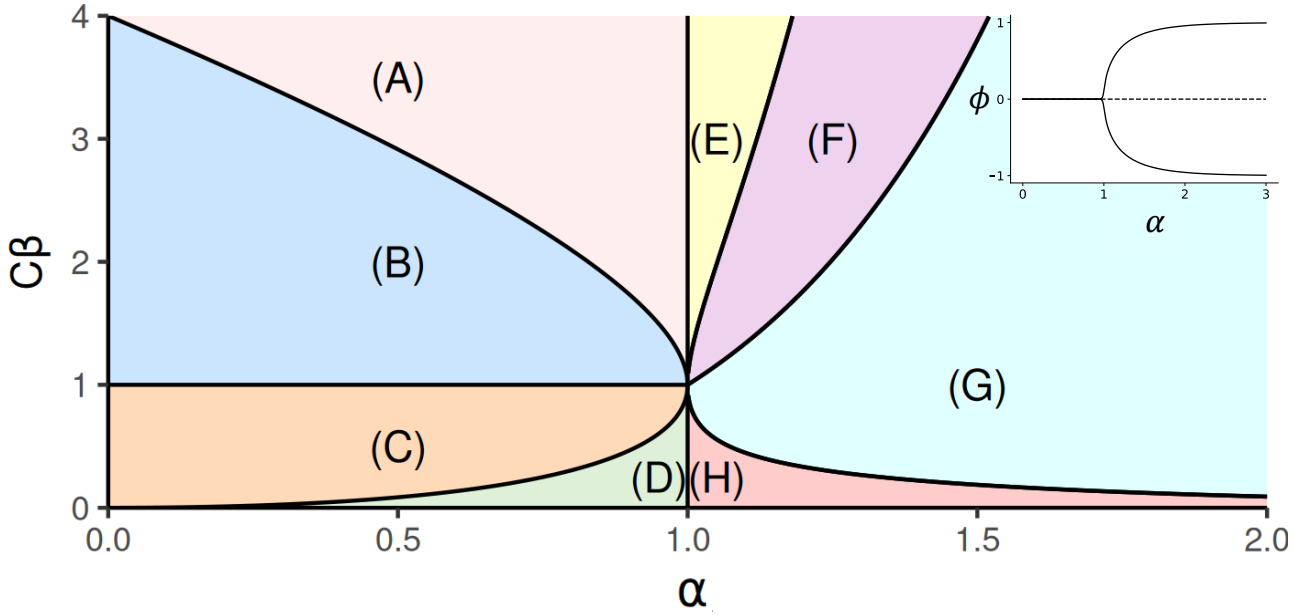


(a) $(0, 0)$ (b) $(\phi^+, 0)$ and $(\phi^-, 0)$

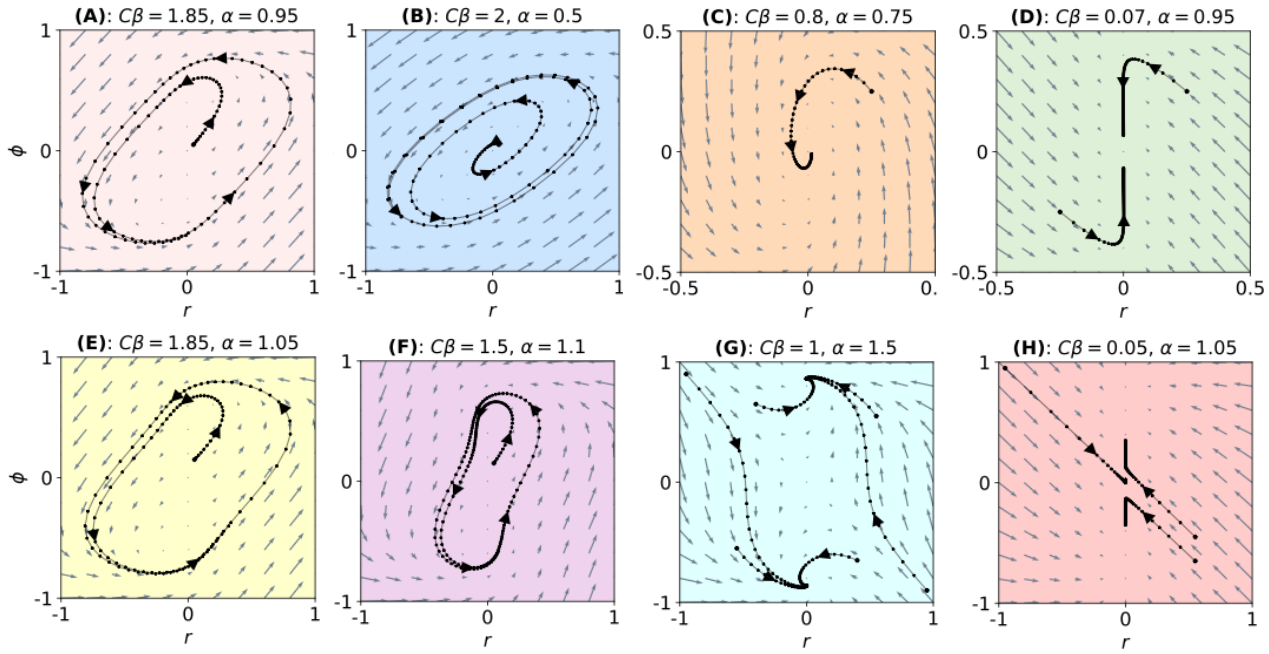
Figure 5.4: **Stability regions of the steady states for Eqs. 5.27-5.28**; the regions for different types of steady states around $(\phi_{ss} = 0, r_{ss} = 0)$ are marked on the left, and those for $(\phi_{ss} = \phi^+, r_{ss} = 0)$ and $(\phi_{ss} = \phi^-, r_{ss} = 0)$ (which coincide due to the symmetry of the sech function), are on the right. The eigenvalues at $(0, 0)$ displayed in Figure 5.4a are: outside the unit circle and real in (A), outside the unit circle and complex in (B), inside the unit circle and complex in (C), inside the unit circle and real in (D), one inside and one outside the unit circle in (I). They are found by evaluating equations 5.29, 5.30 for different parameter values. The eigenvalues at $(\phi, 0)$ are: outside the unit circle and real in (E), outside the unit circle and with an imaginary component in (F), inside the unit circle and with an imaginary component in (G), inside the unit circle and real in (H). They are found by evaluating equations 5.31, 5.32 for different parameter values.

Figure 5.4 is drawn by solving numerically for the points $C\beta$ for which the eigenvalues in equations 5.29, 5.30, on the left, and equations 5.31, 5.32, on the right, switch between the different stability regions. Specifically, we solve for when the complex part of the eigenvalues are equal to zero or not equal to zero, and the eigenvalues are within or outside the

unit circle.[1]



(a) Stability regions for returns and sentiments: in the region where $\alpha < 1$, we observe one steady state where $r_{ss} = 0$ and $\phi_{ss} = 0$; in the region where $\alpha > 1$, we observe three steady state where $r_{ss} = 0$ and ϕ_{ss} can take on values of ϕ^+, 0, ϕ^-. As noted in the inset bifurcation diagram (upper right) the ϕ^+ and ϕ^- are stable steady states and are indicated with a solid line, whereas $\phi_{ss} = 0$ is unstable and is indicated by a dotted line.



(b) Phase diagrams with different parameters: this plot shows the different trajectories that sentiments and returns can take when the system begins outside of a steady state.

Figure 5.5: **Stability of Returns and Prices;** The behaviour of price returns changes qualitatively with the intensity of social contagion (α) and the financial capacity of hype investors ($C\beta$).

[1] It is important to note that we do not observe certain system behaviours within our parameter space for α, β, C - chosen to be consistent with social contagion and trend-following behaviours discussed in the previous chapter. We do not observe eigenvalues crossing outside of the unit circle at -1; this would only be possible if we considered negative parameter values. We also only observe x_1 crossing outside and x_2 remaining inside the unit circle at $\alpha = 1$; a state where the system transitions between stability regions.

Common system behaviours Table 5.1 shows some types of common behaviours around steady states, as they relate to the Jacobian eigenvalues. The key behaviours of relevance are: (i) the existence of an imaginary part of an eigenvalue, and (ii) whether the modulus of an eigenvalue lies within the unit circle. We explore eigenvalues in parameter regions that are consistent with our empirical analysis in the previous chapter: namely where $C\beta > 1$ and $\alpha > 1$. For these parameter values, the eigenvalues cannot be negative – therefore, we explore when the modulus becomes greater than one.

Table 5.1: System behaviour around steady states and Jacobian Eigenvalues

$\|x_i\| < 1$	$\mathbb{I}(x_i) = 0$	stable node / sink
$\|x_i\| < 1$	$\mathbb{I}(x_i) \neq 0$	stable focus / spiral sink
$\|x_i\| > 1$	$\mathbb{I}(x_i) = 0$	unstable node / source
$\|x_i\| > 1$	$\mathbb{I}(x_i) \neq 0$	unstable focus / spiral source

Notes: The behaviour of our system around our steady states is dependent on the eigenvalues of the Jacobian at the steady states. Here we describe some of the most common eigenvalue combinations and behaviours. For a more comprehensive review, see Hommes (2013).

Figure 5.5a shows the different steady states that emerge within our system from the Jacobian for different parameter values of α and $C\beta$. Qualitatively different behaviours emerge within the different stability regions of Figure 5.5a. These are illustrated in Figure 5.5b, where the paths display cyclicality all regions except **(D)** and **(H)**, quasi-periodic dynamics in regions **(A)**, **(B)**, **(E)**, **(F)**, or direct convergence in regions **(D)**, **(H)**.

Overview of Stability Analysis The system exhibits different types of behaviour, depending on parameter values α, β, C. We highlight key system dynamics, as they relate to Figure 5.5a, below.

- **Pitchfork Bifurcation**: As parameter α becomes greater than one, we observe a supercritical pitchfork bifurcation in ϕ. When $\alpha < 1$, the system has one stable steady state for the sentiment variable $\phi = 0$. When $\alpha > 1$, the system has one unstable steady state at $\phi = 0$, and two stable steady states: one above zero and one below zero. This

indicates that in an environment with strong social reinforcement ($\alpha > 0$), individual sentiments can persist, even in the absence of any change in returns - this analytically underscores the conditions for animal spirits or long-term memory in the social setting. The pitchfork bifurcation is displayed in the inset in Figure 5.5a. Regions **A**, **B**, **C**, **D** are those where ϕ exhibits a single steady state; regions **E**, **F**, **G**, **H** are those where ϕ exhibits two stable steady states and one unstable steady state.

- **Stable versus Unstable System**: We consider whether the modulus of the system's eigenvalues moves outside of the unit circle. Specifically, from the functional form of the eigenvalues, one can observe that the eigenvalues do not become negative for our parameters of interest. For this reason, we consider when the modulus of the eigenvalues becomes greater than one. Analytically, we observe that the eigenvalues are smaller than one in the regions **C**, **D**, **H**, **G**: in these regions, we observe that returns and sentiments converge to their steady states. The system has eigenvalues that are larger than one in the regions **A**, **B**, **E**, **F**: in these regions, we observe the system orbiting the steady state(s) (later analysis of the Lyapunov exponent allows us to study the long-term behaviour of these orbits). Such system behaviour can help explain well-established economic tendencies, such as mean-reversion in the markets and large asset returns in the absence of fundamental news.

- **Oscillatory Behaviour**: the existence of an imaginary part of an eigenvalue determines whether the system exhibits oscillatory behaviour. The system has a non-zero imaginary part of the eigenvalues in regions **B**, **C**, **F**, **G**. The impact of oscillatory behaviour is best seen through the comparison of the system's behaviour in regions with, versus without, the an imaginary eigenvalue in Figure 5.5b. In the trajectories for price and sentiment in Figures **D**, **H**, both variables converge directly to their steady states. In the direct counterpart to these figures for the system with imaginary eigenvalues (Figures

C, **G**), on the other hand, returns and sentiments fluctuate along their trajectories before converging to a steady state. Similarly, in regions **A** and **E** where eigenvalues have a zero imaginary part, the system directly moves to cycle around the steady state(s). In regions **B**, **F** minor oscillations along the trajectory are visible (the diversion is relatively small, as it is our objective to explore the state-space in regions where parameter values are consistent with our empirical observations from the previous chapter).

5.3.2 The recent impact of WSB

What has happened to the markets with the influx of eager retail traders onto investment-centric discussion forums? Despite investor discussion platforms having some traction from the early 2000s (with niche groups of investors using myForexBook, StockTwits, or SeekingAlpha), the rise to prominence of the WSB subreddit forum constituted the first widespread coordination attempt among retail traders. By the end of 2023, the forum had gained 15M followers.[2] Shortly after the GameStop short squeeze, news outlets noted the shift in the investment climate, such as the fact that one in five investors claimed that they had used Reddit to drive an investment decision.[3]

In this section, we study the effects of changing the key parameters in our dynamic system at various points in time and for different assets discussed on WSB. Several parameters we expect to remain relatively stable over time and are, therefore, of low interest to this chapter. These are not driven by social media, but rather characteristics of the investor population. Using survey data, we assume that the average hype investor holds $10,000 in funds.[4,5] We estimate λ to be approximately one.

We are able to observe α and β for users active on WSB, as well as for specific stocks. We

[2] https://subredditstats.com/r/wallstreetbets

[3] https://markets.businessinsider.com/news/stocks/reddit-retail-investor-trend-survey-social-media-wallstreetbets-wsb-stock-2021-4-1030366600

[4] https://andriymulyar.com/blog/how-a-subreddit-made-millions-from-covid19

[5] https://www.reddit.com/r/wallstreetbets/comments/a2qvj3/wsb_demographics_survey_results_a_portrait_of/

take into consideration the evolution in the number of users N discussing an asset as a function of returns and sentiments, endogenously increasing capacity C. An adapted epidemic contagion model serves particularly well in predicting the number of new discussants for an asset, based on the number of users previously discussing it and its returns. These findings are detailed in Appendix 5.C. Variation in user count compels us to vary C, and gauge what dynamics dominate market stability using our model.

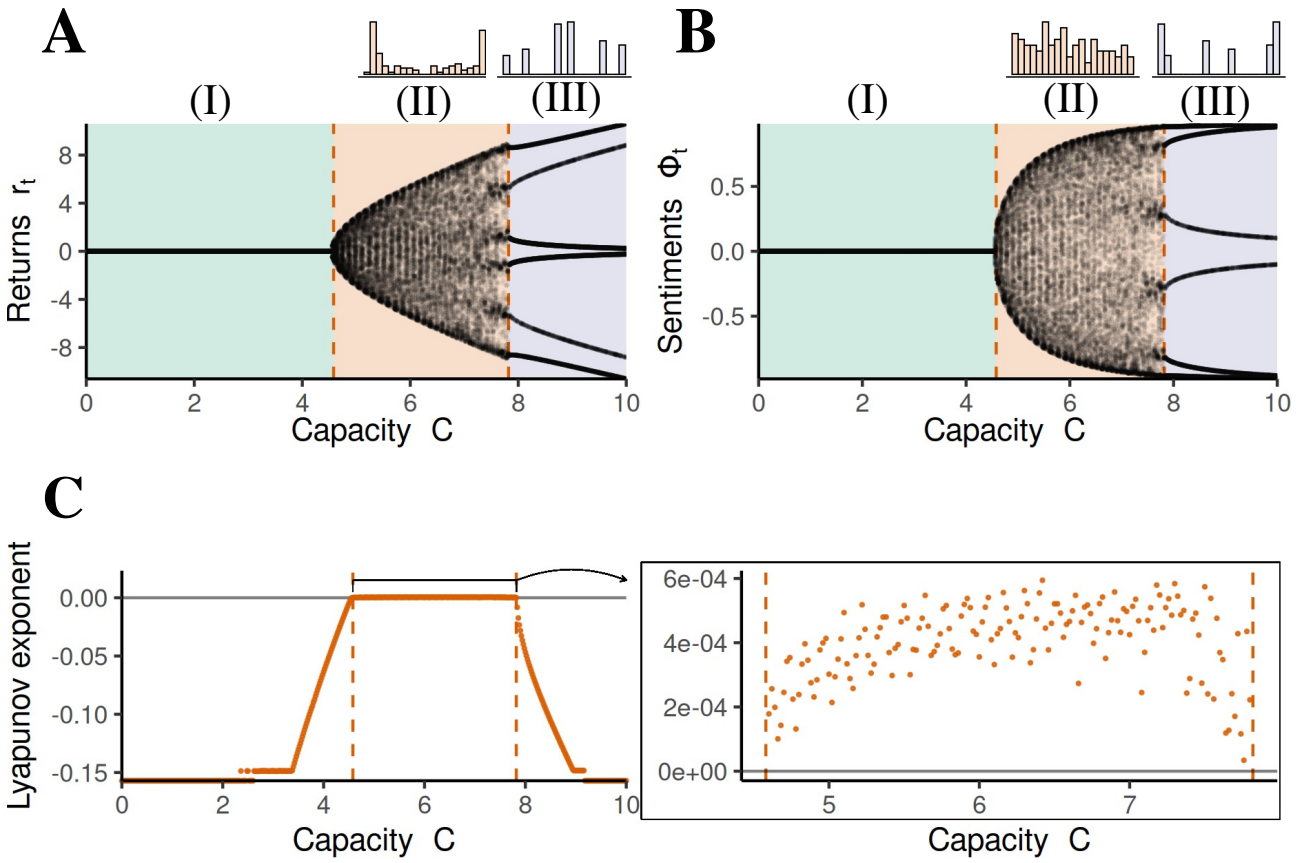


Figure 5.6: **Sentiment and Return Bifurcation Diagrams**; The final value from the 1000^{th} timestep of the dynamic system in Eqs. 5.27-5.28 are plotted for the average α and β values observed on WSB ($\alpha = 0.05$ and $\beta = 0.22$) with different values for C in plots **A** and **B**. Plot **A** displays final values for returns, while **B** displays sentiments. Capacity C varies between zero and ten. For each C value, we perform 100 simulations with initial values for r drawn from a normal distribution with mean zero and standard deviation of 0.05, and with ϕ initiated at zero. We plot the distribution of the returns and sentiments for all 100 simulations at the 1000^{th} time step for $C = 6$ above regions **(II)** and $C = 8$ above region **(III)**. Plot **C** displays the highest Lyapunov exponent estimated using the Jacobian Matrix algorithm, as described in Wu & Baleanu (2015).

Typical system behaviour Across all assets in our sample, we find an average value of $\alpha = 0.05$ and $\beta = 0.22$ (both coefficients are significant at the 1% level) – the estimates differ slightly to those in Chapter 4 as we use our complete sample of data to perform the esti-

mation as our explicit goal for this model is to study tipping points (so capturing the GME short squeeze in our sample is of particular relevance). Figure 5.6 displays the resulting price and sentiment dynamics that occur as we vary C. For each value of C, we display our the final returns r_t and sentiments ϕ_t from 100 different simulations at the 1000^{th} time step. The simulations are initialized with r_0 drawn from a normal distribution with mean zero and standard deviation of 0.05. The goal is to better understand how a small perturbation from the know steady-state of $r = 0$ and $\phi = 0$ may impact price and sentiments when social investors are present.

Panels **A** and **B** in Figure 5.6 demonstrates that capacity C would have to be unreasonably large – total funds of WSB users must be over four times greater than the asset's market capitalization – for the system to fall out of stable region **(I)**, and into a regime with quasi-periodic cycles **(II)**. Despite such a capacity being high for the market as a whole, however, for specific stocks such as GME, RAD and AMD with lower market caps – reaching a capacity that is several times the market cap of these stocks (using social media as a coordination platform) is well within the realms of possibility. These parameters reside largely in regions **(C)** and **(B)** from Figure 5.5.

Figure 5.6's region **(II)** displays quasi-periodic dynamics. The system's behavior does not precisely repeat itself, and the return time-series appears irregular. This pattern is evidenced by an estimated Lyapunov exponent close to zero, as seen in Panel **C**. However, the concentration of final values of r_{1000} and ϕ_{1000} demonstrates a higher probability for us to observe extreme values for r_t and ϕ_t (a uniform distribution for r_{1000} and ϕ_{1000} would have suggested chaotic dynamics). The degree of periodicity (as well as the maximum and minimum values for the fluctuations) in this region vary for different values of C.

In region **(III)** of Figure 5.6, the system enters a stable cycle. This is evidenced by the estimated Lyapunov exponent switching from zero, to a range of negative values in Panel **C**.

The system moves between a finite number of states, without visiting intermediate values for r and ϕ. The contrast between the irregular cycles in region **(II)** and the regular kind in **(III)** is discussed below.

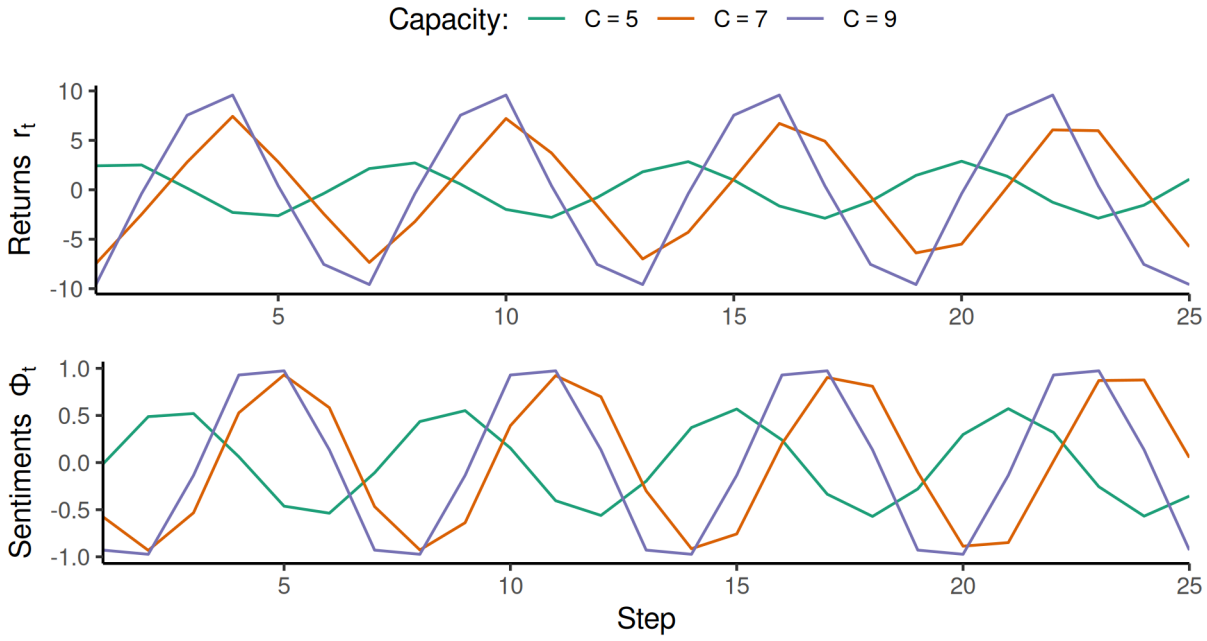


Figure 5.7: **Increasing the capacity C of hype investors introduces quasi-periodic cycles, and eventually fully periodic cycles into the dynamics of returns and sentiments**;

Periodic dynamics Figure 5.7 helps understand the dynamics behind the bifurcations observed in Figure 5.6. For three values of capacity C, we let the simulation run for 1000 iterations. At that point, the dynamics stabilise, and exhibit somewhat irregular cycles when capacity is five and seven. At those values, the system exhibit a Lyapunov exponent equal to zero, meaning that two series outside of fixed point equilibria do not converge, nor diverge, at any rate. Once capacity is large enough, for example at nine times the valuation of the asset, the system exhibits periodic cycles, and a negative Lyapunov exponent.

System behaviour for specific assets Have hype investors actually destabilized the markets? From our average estimates, it is clear that WSB users would have to have access to vast amounts of capital to cause market fluctuations. However, the picture is more nuanced

on a granular level. To demonstrate this heterogeneity, we estimate α and β for specific stocks.

We present a sample of our estimates in Table 5.2. Within our sample, we observe that WSB users exhibit varied degrees of trend following and social contagion. In the case of AAPL, for example, trend following appears to play a significant role. For GME, both trend following and social contagion have modest, but persistent effect. The outsized market capitalization of AAPL relative to GME, as well as the number of active discussants, played an important difference.

Table 5.2: Estimates for α and β for different assets

Ticker	Name	Market Cap. (Dec. 2020, $M)	α	β
VXRT	Vaxart, Inc.	837	2.05(0.21)***	0.77(3.62)
TLRY	Tilray Brands, Inc.	1,078	0.35(0.07)***	0.38(0.70)
PCG	PG&E Corp.	25,184	0.29(0.10)***	2.35(2.03)
GME	GameStop Corp.	1,030	0.07(0.01)***	0.13(0.04)***
EA	Electronic Arts, Inc.	36,881	0.42(0.12)***	27.78(10.51)***
AMD	Advanced Micro Devices, Inc.	111,407	0.09(0.03)***	1.62(0.91)*
AMC	AMC Entertainment Holdings, Inc.	355	0.07(0.02)***	-0.08(0.14)
AAPL	Apple, Inc.	2,086,461	0.12(0.04)***	9.44(3.40)***
MU	Micron Technology, Inc.	74,922	0.01(0.72)	4.95(0.00)***
RAD	Rite Aid Corp.	730	0.06(0.47)	3.40(0.01)***

Notes: this table lists the parameter estimates for α and β for specific stocks in our WSB sample. Standard errors (calculated while accounting for autocorrelation with one lag and heteroskedasticity) are presented in parentheses next to the estimate. For implementation please reference Seabold & Perktold (2010).
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Figure 5.8 demonstrates how specific values for α and β can lead to substantially different dynamics. In the case of AAPL and EA, for example, the combination of a high degree of trend-following and social contagion results in a bifurcation occurring when social investors possess much lower capacity. In the case of EA, we observe destabilizing impacts when the total capacity C of social investors is less than 5% of the stock's market cap – a quantity not beyond the realm of imagination, given that WSB reached 15M followers by 2023.

In the case of GME, the lower estimates for α and β imply stable asset price dynamics

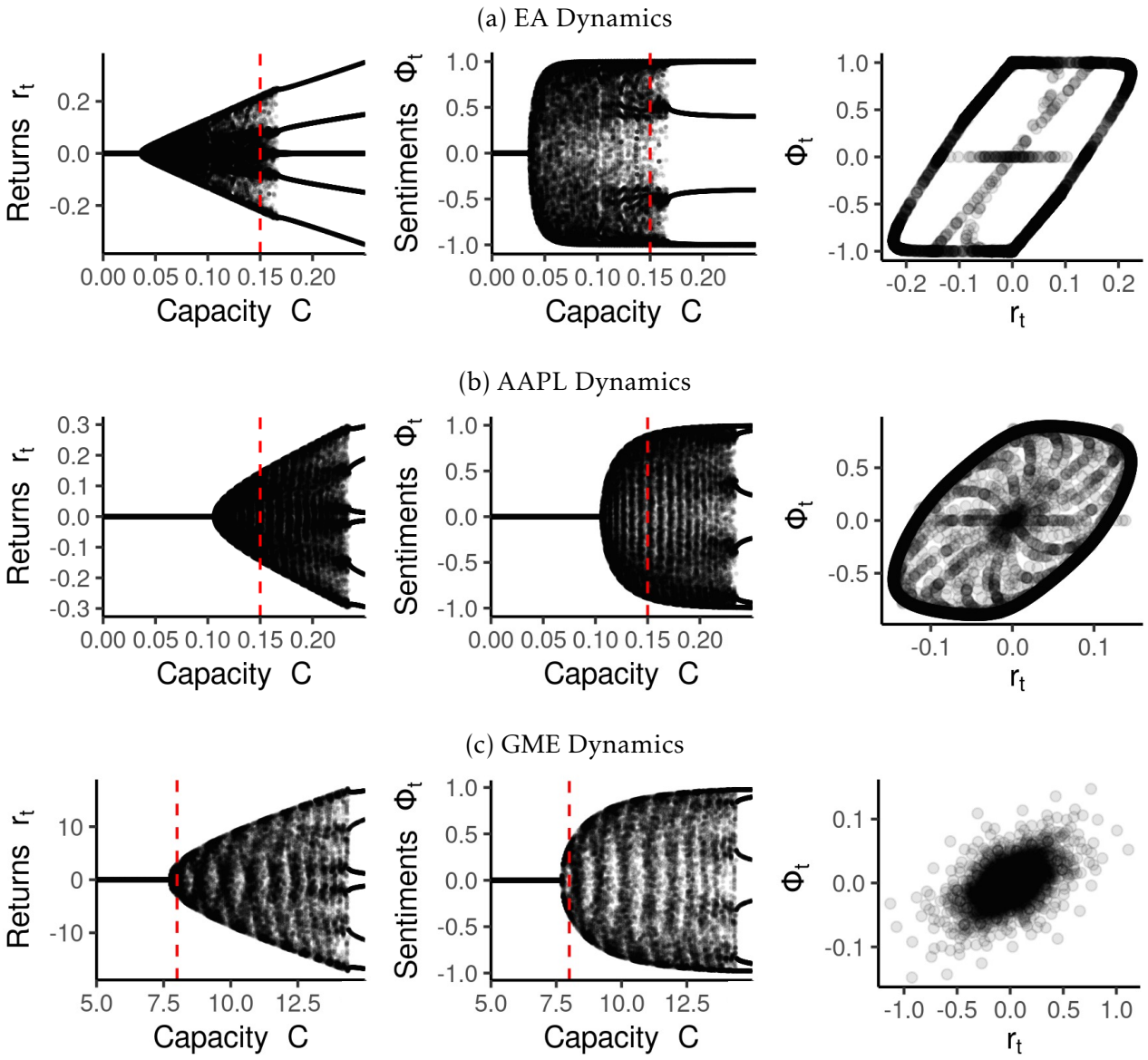


Figure 5.8: **Specific Stock Returns and Sentiments Entering Unstable Territory**

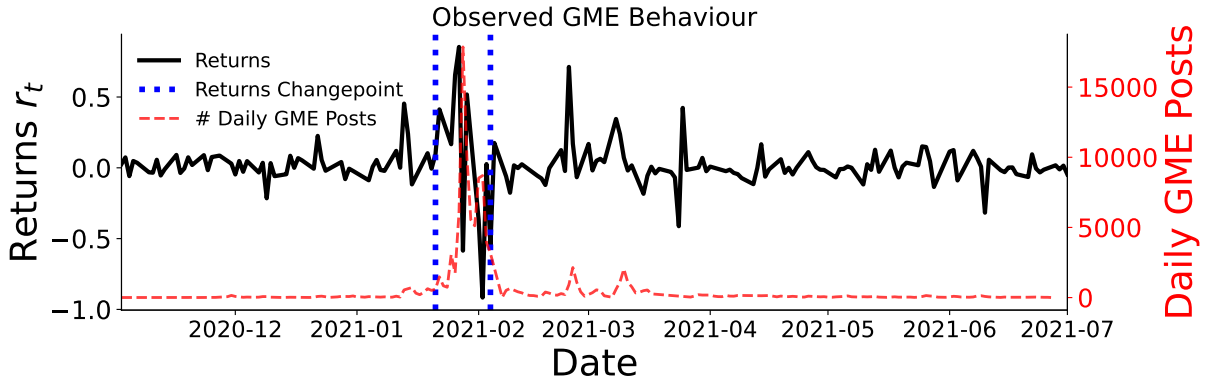


Figure 5.9: **GameStop Asset Returns and WallStreetBets Interest**

until hype investors reached a certain capacity. However, after a certain capacity threshold is exceeded for hype investors, the asset returns and investor sentiments undergo a bifur-

cation. This hypothesis is consistent with the data, where GME returns appeared relatively stable until the WSB discussion forum increased in size and focused it's attention on GME. At this point, the time-series of returns experienced a changepoint, as displayed in the bottom of Figure 5.9. These returns again reverted to a more stable regime once hype investors lost interest in the asset, and their overall capacity diminished.

Appendix

5.A Equilibrium Model Appendix

Complementarity in demand We study the role of complementary investment decisions. To that end, the model operates in two stages. In the first stage, investors build their expectation for the asset's value, using observed signals from their peers and their expectation of the market-clearing price as a function of the expected, and as of yet hidden, supply shock. In the second stage, the asset supply shock is revealed, and investors execute their trades according to their demand curve.

Investor i who expects value $\mathbb{E}_i(v)$ and understands the price setting mechanism. Investor i's maximised payoff from Eq. 5.1 is

$$\mathcal{L}(\phi_i^*) = \frac{\mathbb{E}_i^2(v-p)}{2\gamma\mathbb{E}_i(v-p)^2} \tag{5.33}$$

$$= \frac{1}{2}\mathbb{E}_i(v-p)\phi_i^* \tag{5.34}$$

$$= \frac{1}{2}[\mathbb{E}_i(v)-\mathbb{E}(v)]\phi_i^* + \frac{\gamma\sigma^2}{2}\left(\frac{1}{N}\sum_j^N \phi_j^*\right)\phi_i^*, \tag{5.35}$$

where $\mathbb{E}(v) = 1/N\sum\mathbb{E}_i(v)$ as before. Here, investors base their price expectations on the simple equilibrium in Eq. 5.3, but use their personal expectations and constant uncertainty σ^2 to forecast price in the second stage. Eq. 5.35 demonstrates that the investor's payoff depends on their peers in two regards. First, payoffs increase to the degree that the investor in question expects to outperform others, in terms of the value they realise in the asset. This is seen in the first component, by which buying(selling) the asset increases the payoff to the extent that i's expected value $\mathbb{E}_i(v)$ is higher(lower) than that of their peers. Second, the payoff increase by the average optimal asset demand of all investors in the economy.

The asset demand model predicts that social interactions – knowledge of other's asset purchases – is a significant component of investors' welfare *in expectation*. Eq. 5.35 is a well-known formulation for strategic interactions between agents acting under quadratic loss (Zenou 2016). Deriving Eq. 5.35 with respect to two investors' demands reveals their strategic complementarity:

$$\frac{d^2\mathcal{L}(\phi_i^*)}{d\phi_j^* d\phi_i^*} = \frac{\gamma\sigma^2}{2N} > 0. \tag{5.36}$$

The emergence of strategic complementarities is due to a crowding effect that investors have on price. The higher asset demand by other investors, the higher the realised price will turn out to be. Before the value of the asset is revealed, investors are motivated to gauge demand by others to better estimate what the price will be, in excess of their personal valuation.

The unweighted average of peer sentiment in Eq. 5.35 emerges because we did not provide a specific mechanism by which information about asset demand is transmitted. The acquisition of information under some cost to the investor is an interesting extension, although already studied by Hellwig & Veldkamp (2009). Their study offers more rigorous insight into the manifestation of strategic complementarities, as well as the emergence of multiple equilibria, when investors seek to learn about the underlying price from a set of possible signals.

5.B Complex Systems Model Appendix

Asymmetric Upside and Downside Probabilities A potential shortcoming of our approach is the fact that $U(d_{i,t+1} = +1)$ and $U(d_{i,t+1} = -1)$ are symmetric in parameters α, β and γ. However, it is well-documented that individuals are not symmetric in their perception of downside loss versus upside potential (a tendency referred to as 'risk aversion'). On

the other hand, some may argue that 'hype investors' are not representative and may have risk-seeking tendencies. These arguments potentially motivate keeping the parameters for upside potential and downside loss distinct. This analysis allows us to arrive at the quantal response function for the probability of positive or negative sentiment:

$$\phi^{D}_{i,t+1} = \begin{cases} +1, & \text{with probability} \dfrac{\exp\left[(\alpha\phi_t + \beta r_t - \gamma r_t^2)/2\lambda\right]}{\exp\left[(\alpha\phi_t + \beta r_t - \gamma r_t^2)/2\lambda\right] + \exp\left[(-\alpha\phi_t - \beta r_t - \gamma r_t^2)/2\lambda\right]}, \\ -1, & \text{with probability } \dfrac{\exp\left[(-\alpha\phi_t - \beta r_t - \gamma r_t^2)/2\lambda\right]}{\exp\left[(\alpha\phi_t + \beta r_t - \gamma r_t^2)/2\lambda\right] + \exp\left[(-\alpha\phi_t - \beta r_t - \gamma r_t^2)/2\lambda\right]}. \end{cases}$$

We briefly explore this asymmetry in our empirical results, however, accurately estimating such parameters separately is potentially a fruitful extension of this work.

5.C Hype investor interest

In the WSB context, we would expect awareness about specific assets to spread from one user to another, in line with the observations of Shiller (2005), Banerjee (1993) and Banerjee et al. (2013). The emphasis of this section is not on identifying a causal relationship, but rather understanding the dynamics which govern asset interest among investors. These insights, combined with a mechanism for investors' joint sentiment adoption, allow us to paint a more complete picture of retail investor decision-making and the resultant stock market dynamics.

We model the log-odds of an author posting about stock j over a baseline using the following linear model:

$$l(a_{j,t}) = \log\left(\frac{a_{j,t}}{s_t}\right) = c a_{j,t-1}(1 - a_{j,t-1}) + d a_{j,t-1} + \beta_1 \bar{r}_{j,t-1} + \beta_2 \sigma^2_{j,t-1} + X_j \beta_4 + \zeta_{j,t}, \tag{5.37}$$

where t denotes time (in weeks), the baseline s_t is the probability of posting about a stock that is not widely discussed within the forum (a stock that is mentioned in fewer than 31

submissions within our sample), $a_{j,t-1}$ is the share of all active investors who post about ticker j at times $t-1$ ($a_{j,t} \in [0,1]$ for all j and t), $\bar{r}_{j,t-1}$ is the average log-return in $t-1$, and $\sigma^2_{j,t-1}$ is the variance of the same log-returns. X_j is a vector of stock dummies.

Our framework is inspired by Banerjee et al. (2013) – individuals become interested in an asset due to their peers and thanks to a public signal of the asset's performance. Parameter c captures the rate of independent mixing between investors aware of stock j, $a_{j,t-1}$, with unaware investors, $1 - a_{j,t-1}$. Parameter d captures the rate at which aware investors lose interest. The latter terms control for the asset's perceived profitability and riskiness. Parameter β_1 is a 'quality of signal' term capturing how well the asset has performed in the past, and β_2 a 'noise of signal' term, measuring the asset's recent volatility. We propose that coefficients c and β_1 are positive – implying that these dynamics contribute to increased interest in a stock – while d and β_2 are negative.

The choice to aggregate over weeks is done to address the sparsity of submissions, especially pre-2017. In addition, we categorise stocks mentioned fewer than 31 times since January 2012 into an 'other stocks' group, which forms our benchmark s_t.

We also consider a different formulation where we test for the direct impact of historical peer sentiments and the interactions between historical sentiments and returns / volatility: $\bar{\phi}_{j,t-2}\bar{r}_{j,t-1}$, $\bar{\phi}_{j,t-2}$ and $\bar{\phi}_{j,t-2}\sigma^2_{j,t-1}$. This formulation allows us to evaluate whether WSB users are more likely to discuss a stock if the predictions of their peers have been correct, and accurate, in the past.

5.C.1 Results

The OLS estimates of our model in Eq. 5.37, presented in Table 5.C.1, demonstrate that WSB users follow each other in their choice of stocks. There is strong evidence that the homogeneous mixing property partially explains the uptake of new assets: using estimates

Table 5.C.1: Stocks discussed on WSB

	Dependent variable: $l(a_{j,t})$			
	(1)	(2)	(3)	(4)
$a_{j,t-1}(1-a_{j,t-1})$	71.58*** (5.62)	71.52*** (6.48)	41.32*** (7.63)	40.45*** (8.34)
$a_{j,t-1}$	−30.20*** (4.65)	−29.99*** (5.32)	−15.19** (6.04)	−14.75** (6.62)
$\bar{r}_{j,t-1}$	0.92*** (0.29)		1.83*** (0.26)	
$\sigma^2_{j,t-1}$	−1.52*** (0.30)		0.25 (0.36)	
$\bar{\phi}_{j,t-2}\bar{r}_{j,t-1}$		2.09** (0.83)		3.55*** (0.68)
$\bar{\phi}^2_{j,t-2}\sigma^2_{j,t-1}$		−4.69*** (1.77)		2.14 (1.61)
Constant	−4.02*** (0.01)	−4.04*** (0.01)		
Ticker FE	No	No	Yes	Yes
Number of obs.	25,637	12,708	25,637	12,708
Adjusted R^2	0.25	0.35	0.10	0.14
F-statistic	2,166.05	1,699.32	825.40	660.10

Notes: this table presents OLS estimates for the log-odds of users discussing stock j in week t, over a collection of stocks that are mentioned fewer than 31 times. Explanatory variables include: the lag in the share of authors discussing j, $a_{j,t-1}$, the interaction with the share of authors not discussing j, $a_{j,t-1}(1-a_{j,t-1})$, as well as the lag in stock j's weekly average log-return, $\bar{r}_{j,t-1}$, and variance, $\sigma^2_{j,t-1}$. In columns (2) and (4), the average log-return is multiplied by the two period lag in the average sentiment expressed among WSB submissions on stock j, $\bar{\phi}_{j,t-2}$, and the variance in log-returns by the same sentiment's square, $\bar{\phi}^2_{j,t-2}$. Columns (3) and (4) include stock-specific fixed effects. Accompanying standard errors, displayed in brackets, are clustered at the stock level, and calculated in the manner of MacKinnon & White (1985).
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

in column (1), an increase in the share of authors discussing stock j from 0.1 to 0.2 increases the ratio of authors discussing j over 'other stocks' in the following week by approximately threefold. This is contrasted by an increase from 0.2 to 0.3, which prompts a decline in the ratio of authors discussing j over 'other stocks' by 50% – the difference is driven by the large negative coefficient on $a_{j,t-1}$. This is strongly reminiscent of epidemic contagion models, adapted to the spread of narratives Banerjee (1993), Shiller (2017).

When we consider the impacts of stock-specific variables in isolation, presented in columns (1), (3) in Table 5.C.1, volatility and returns appear to be leading factors for authors deciding what asset to discuss. Average, historical returns are statistically significant at the 1% level in columns (1) and (3), indicating that discussion sizes are stimulated by large, notably positive, returns. Examining the coefficient in column (3), a stock that experienced a 5% greater return in one week is the subject of about 7% more submissions than usual. Volatil-

ity appears to play a greater role in our formulation without ticker-specific effects, with its significance declining from column (1) to (3) – a factor perhaps explained by the choice of hype investors to overlook recent volatility in certain assets, but not others. Our alternative formulation presented in columns (2) and (4), estimating the effect of the correctness and consistency of past WSB predictions in an asset, appears to have some significance in explaining asset interest.

Chapter 6: Has Social Investing Destabilized Financial Markets?

Chapter 4 documents two important contributors to sentiment formation among social investors – extrapolation and peer effects. Chapter 5 considers potential emergent market dynamics that can result from investors operating based on their sentiments (driven by the previously documented peer effects and extrapolation). This chapter reconciles our modeling exercises with observations from the market. We provide evidence: of the fact that investors trade based on their expressed sentiments (using position screenshots) and expressed asset interest (using Trade and Quotes data), of the market impact of viral online content, and of the participation of social investors in assets experiencing bubbles.

We investigate changes in retail trader asset demand by tracking retail investor trades, using the methodology from Boehmer et al. (2021). We observe that changes in asset interest on WSB explain a significant fraction of changes to retail trading behaviour (Section 6.2).

A key challenge to analysing returns is that sentiments and returns are co-determined in equilibrium. If returns are high, sentiments are also high. Conversely, if sentiments are high, buying pressure will also increase returns contemporaneously. Our strategy exploits the heavy tails in the popularity of discussions on WSB for identification and demonstrates that sentiments expressed in submissions move asset prices as a function of their virality – as proxied by their comment count (Section 6.3).

In a final empirical test, we identify bubble-like dynamics in assets and verify that WSB users have a significantly greater interest in assets whose price experiences as a sharp rise, but subsequently implodes, as compared to those that have a sustained increased in price follows by price stability (Section 6.4).

Additional work not included in the thesis also demonstrates reversion in returns and asset price cyclicality in the presence of social investors.

6.1 Isn't all of this just talk?

Why should we care about the sentiments people express about assets online? Anecdotally, the GameStop short squeeze demonstrated that the online discussions on the WSB forum have impact on assets. However, this does not constitute evidence of the fact that people follow through on the investment strategies they discuss online *systematically*.

To address the concern that WSB sentiment data has limited impact on investment decisions, we utilize screenshots that users post of their investment positions to test whether they follow through on their expressed sentiments. We extract approximately 9,000 images from WSB – we focus only on image-related URLs (such as ones with the domain name 'imgur', an image-hosting site) mentioned in posts of authors who had previously posted about a ticker. We hand-annotate a third of the images. Specifically, we manually annotate the image, if it is a position screenshot, with i) the tickers in the screenshot and ii) the positions (long or short) the author displays. The position taken by author i in asset j, $B_{i,j}$, is annotated as +1 if the author is long in the asset, and -1 if the author is short.

We note that the sample of screenshots is biased. Authors on WSB are socially incentivized to share extreme losses (known at 'loss porn' on the forum) or gains. We, therefore, observe relatively few positions, as compared to sentiments. The positions data is also skewed towards long positions, which is consistent with the skew towards bullish sentiments on the forum. However, despite these shortcomings, the positions provide sufficient variety in investment strategies to test whether people trade on their expressed sentiments on WSB.

We match the ticker screenshot to a submission posted before/simultaneously with that screenshot by the same author and about the same ticker. We regress the most recently expressed sentiment by author i about asset j (our key sentiment variable $\Phi_{i,j}$) on the log-

odds of the position $B_{i,j}$ extracted from their screenshot being long versus short:

$$\log\left(\frac{\mathrm{P}(B_{i,j}=+1)}{\mathrm{P}(B_{i,j}=-1)}\right) = \lambda^s \Phi_{i,j} + u^p_{i,j,t}, \tag{6.1}$$

where λ^s measures the pass-through rate of sentiment into eventual investment positions. In an alternative formulation, we represent the past sentiment as three categorical variables: $\phi^{-1}_{i,j}$, $\phi^{0}_{i,j}$, $\phi^{+1}_{i,j}$, which take on a value of one if the author's sentiment is labeled short, neutral or long, respectively (and a value of zero otherwise).

Table 6.1: Follow-through on WSB Advice

	Dependent variable: Position in Asset j of Author i	
	$B_{i,j}$ - categorical	
	(1)	(2)
$\Phi_{i,j}$	0.72 *(0.09)* ***	
$\phi^{-1}_{i,j}$		-0.97 *(0.29)* ***
$\phi^{0}_{i,j}$		0.66 *(0.21)* ***
$\phi^{+1}_{i,j}$		1.84 *(0.27)* ***
Observations	278	278
Pseudo-R^2	0.14	0.17

Notes: This table presents estimated log-odds coefficients for two logit models for the relationship between the sentiment expressed by user i about asset j and the subsequent long/short position the user reports (see Eq. 6.1). Sentiment estimates are presented in two ways: (1) the continuous log-odds of the author expressing positive over negative sentiment $\Phi_{i,j}$, and (2) as a categorical variable where $\phi^{-1}_{i,j}$ corresponds to the expression of negative sentiment, $\phi^{0}_{i,j}$ - neutral, $\phi^{+1}_{i,j}$ - positive.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Results Table 6.1 presents the coefficients of past sentiments regressed on future positions, estimated using a logistic regression. We observe that an author's sentiment is highly correlated to their subsequent holdings of the stock. Let us consider the results in column (2) – an author creating a bullish post about an asset raises the probability of a long versus short investment by over six times. Furthermore, this specification of the logistic regression predicts an author's position with over 75% accuracy. The positions data gives us confidence that investors do trade based on their discussions and expressed sentiments. We consider a

more comprehensive sample of retail investor trader in the following section.

6.2 Evidence of trading

We test for a link between discussions on WSB and changes to retail investor demand, approximated by the fraction of retail investor trades executed in the market.

Variable Definition Retail trader activity is identified from Trade and Quote (TAQ) data – a dataset containing all transactions for listed stocks in the United States. We leverage the fact that retail transactions are offered price improvements and, therefore, may execute at a fraction of a penny. To identify retail traders, we first filter trades to those with *exchange code* = 'D' in TAQ. In the remaining trades, we identify those that execute at a fraction of a penny as retail transactions. Specifically, let $Z_{j,t} = mod(100 * P_{j,t}, 1)$, the fraction of a penny associated with the transaction price in stock j at time t. If $Z_{j,t}$ is in the interval (0,0.4) or (0.6,1), the transaction is coded as a retail transaction. We define a metric for retail trade fraction in asset j in week t_w as:

$$F^R_{j,t_w} = V^{\text{retail}}_{j,t_w} / V^{\text{total}}_{j,t_w},$$

where $V^{\text{retail}}_{j,t_w}$ is the sum of the sizes of all trades labeled as retail transactions using the method above in asset j in week t_w, and V^{total}_{j,t_w} is the sum of the sizes of all trades in asset j in week t_w from TAQ. Our variable of interest – the change in retail investor trading fraction – is defined as:

$$\Delta F^R_{j,t_w} = F^R_{j,t_w} - F^R_{j,t_w-1}.$$

We design a similar monthly metric, $\Delta F^R_{j,t_m}$, to track changes in retail trade volumes on a monthly scale. Importantly, a change in news or asset-level characteristics cannot justify a change to $\Delta F^R_{j,t_w}$, since it would affect retail traders and institutional investors. Our metric, therefore, allows us to distinguish between changes to *retail investor preferences* versus overall market shifts.

Our goal is to consider whether discussions on WSB explain variation in the fraction of retail investor transactions in the market. We define a metric tracking the prevalence of discussions about asset j in week t_w on the forum versus the overall number of posts about assets on the forum, F^W_{j,t_w}, defined as the number of posts mentioning asset j in week t_w over the total number of posts mentioning assets in week t_w. Our predictor of interest is defined as:

$$\Delta F^W_{j,t_w} = F^W_{j,t_w} - F^W_{j,t_w-1},$$

the change in the fraction of posts discussing asset j on the forum.

Analytic Approach Our goal is to demonstrate that changes in discussions among retail investors are accompanied by changes in retail trading volumes. We regress change in weekly (monthly) retail trade fractions on the changes in ticker importance on the forum:

$$\Delta F^R_{j,t_w} = \beta_v \Delta F^W_{j,t_w} + \eta_j + \epsilon_{j,t_w}, \tag{6.2}$$

where η_j are ticker fixed effects, and ϵ_{j,t_w} is an error term. We repeat the same exercise, except with variables computed at the monthly time scale, t_m.

Results We propose that tickers that are popular on WSB are more likely to be linked to changes in retail trader order flow. For this reason, we look at the trading patterns of the

Table 6.2: Retail trade volume versus the WSB discussion

	Dependent variable: $\Delta F^R_{j,t}$					
	Weekly Changes			Monthly Changes		
	(1) Pooled	(2) Pooled	(3) Top 25% (market cap)	(4) Pooled	(5) Pooled	(6) Top 25% (market cap)
$\Delta F^W_{j,t}$	0.098***	0.098***	0.167***	0.261***	0.253***	0.505***
	(0.019)	*(0.019)*	*(0.030)*	*(0.049)*	*(0.049)*	(0.033)
Ticker FE	No	Yes	No	No	Yes	No
Observations	3,503	3,503	1,040	776	776	232
R^2_{adj}	0.007	0.000	0.028	0.035	0.011	0.116

Notes: this table presents the OLS estimates for the influence of changes in WSB discussion interests on retail trading patterns. Columns (1), (2), (3) present the weekly estimates, while (4), (5), (6) present the monthly ones. Columns (3) and (6) consider the estimates for the top quartile of stocks, by market cap, within our sample. All standard errors are clustered at the ticker level.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

twenty most popular tickers, by year, on WSB between the years 2017 and 2020. Table 6.2 presents our main result. Changes in discussion popularity of tickers are statistically significant for explaining changes in retail trading behaviour at the weekly and monthly level. The monthly estimates appear more significant and help explain a greater variation in the dependent variables, as per the R^2_{adj}.

In columns (3) and (6), we repeat the exercise but only consider the stocks that are within the highest quartile by average market cap between the years of 2012-2020 within our sample: AAPL, AMZN, BA, BAC, DIS, FB, GE, GILD, MSFT. We observe that the coefficients on $\Delta F^W_{j,t}$ at the monthly and weekly time frame are more significant in this formulation and explain a greater fraction of the variation in changes in retail investor trading activity. At the monthly time horizon, 10% increase in the prevalence of AAPL discussions on WSB is associated with a 5% increase in retail trading activity in the market. Interestingly, the stocks within the top quartile do not include 'meme' stocks - indicating that WSB discussions track retail preferences across a broad spectrum of assets. Including ticker fixed effects

in our model specifications for columns (3), (6) decreases the R^2_{adj}, indicating that the effect cannot be explained by stock-level differences.

Our experiments demonstrate that WSB discussions track changes in retail trading behaviour at the weekly and monthly timescales. These results further justify our modeling approach, highlighting how WSB activity is linked to retail trader demand for assets.

6.3 Evidence of granularity

Our modeling exercise in Chapter 5 shows that heterogeneous opinions can impact asset returns in the presence of heavy-tails in attention. Our empirical exercise in Chapter 4 shows that individual sentiments are not *fully* explained by information from peers or recent returns, as demonstrated by the R^2 in our regressions. There is, therefore, unexplained heterogeneity in individual investor sentiments. An outstanding question is whether these heterogeneous sentiments can survive aggregation across peers and impact asset returns. This section demonstrates that viral content can influence future asset returns; Appendix 6.A offers additional insights into our identification and presents the diversity in influential, viral content identified through our methodology.

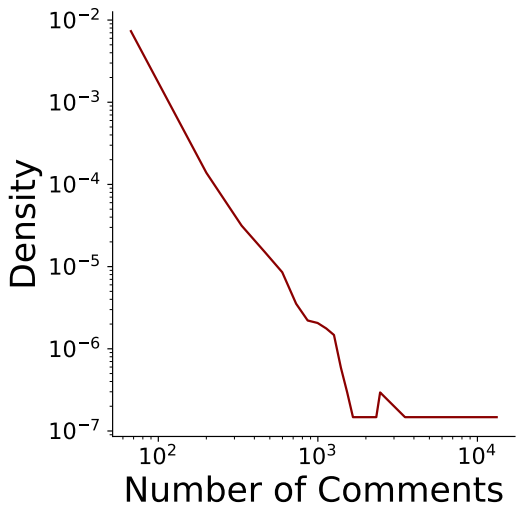
Model framework We remind ourselves of the proposed model framework from the previous chapter (section 5.1), which captures the impact of individual, idiosyncratic demand shocks for an asset by investor i, e_i, on price:

$$p = \sum_{i=1}^{N} s_i \mathbb{E}_i(v) - \gamma \sigma^2 S + \sum_{i=1}^{N} s_i e_i.$$

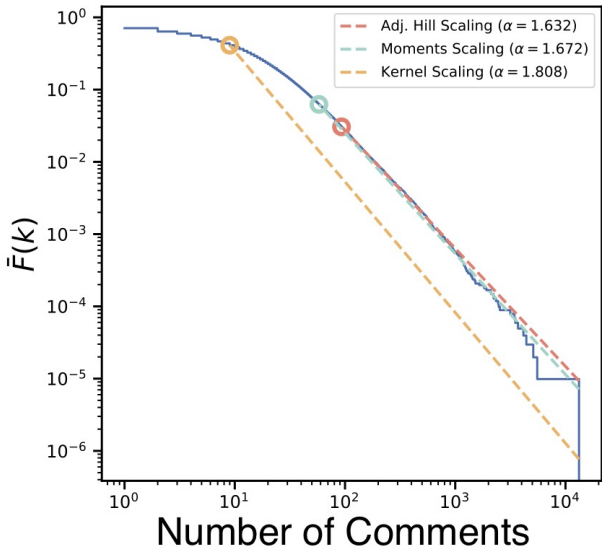
In the absence of 'granularity' among investors, the shocks average out to zero. However, when certain investors are weighted differently to others due to differences in capital or popularity, the shocks can have an impact on price, as highlighted above. If there is a granular

shock at time t, we would therefore expect a change in the log price, which would manifest in a correlation between the granular social shock and returns, as well as increased volatility.

Granularity of social attention We leverage the heavy-tailed structure of WSB discussions for our empirical strategy. The intuition is that certain submissions gather many more followers than others, which we measure using the number of comments they receive. An initial exploration of the data displays the heavy-tail in discussions *between* assets – a handful of tickers are mentioned in thousands of posts, while most assets receive just a small number of mentions. Attention *within* stocks is also dominated by a few, heavily-commented submissions. In our modification to our model, the heavy-tail of attention can result in unexplained variation in sentiments surviving aggregation and impacting returns. We use a Granular Instrumental Variable (GIV) approach to investigate the impact on returns (Gabaix & Koijen 2023, 2021, Galaasen et al. 2020).



(a) Distribution of comments across all submissions



(b) Tail estimate for the distribution of comments

Figure 6.1: **Attention is heavy-tailed**; These graphs show the distribution of comments that submissions that mention a ticker on WSB receive. The left figure plots the distribution on a log-log scale. The right plots the tail exponent, estimated using the three different methods outlined in Voitalov et al. (2019). The tail exponent is estimated to be less than two across all methods – this implies that the tail distribution obeys a power law, and is heavy-tailed.

We begin by establishing that the *distribution of attention* that information shared by

investors online receives is heavy-tailed. We proxy attention by the number of comments that a particular submission receives. Figure 6.1 shows the extreme tail in the distribution of attention – some submissions appear to receive a large following, while the majority are of little interest. We measure the size of the tail in the comment distribution using the approach in Voitalov et al. (2019), who propose several methods for estimating the power-exponent of a distribution's tail – all methods estimate the tail exponent to be less than two, implying that the tail is power-law and heavy-tailed.

Heavy-tailed attention online implies that idiosyncratic information contained within the most popular submissions potentially persist after pooling across all investor's opinions, and may have a disproportionate effect on returns. For our identification strategy, we exploit within-ticker-week variation in attention. Therefore, we filter our sample to weeks and tickers where a sufficient number of submissions are made to distinguish between highly popular and less popular sentiments – we choose five submissions for our cutoff. We subsequently test for whether the week in question exhibits granular social attention by fitting a pareto distribution to the popularity of posts within a week and selecting ticker-week combinations where the exponent is less than two.

The approach has certain shortcomings as there are challenges to finding the exponent with a small sample of data, however, allows us to estimate whether the activity in a ticker in a given week is a good candidate for granular social dynamics. We ensure our result is robust by imposing several different thresholds and varying our approach between applying and not applying the heavy-tail filter; we observe that all results are consistent with the findings presented below.

Preliminary analysis As a preliminary test, we measure the link between granular social activity and volatility. Table 6.3 shows that there is a statistically significant link between the standard deviation of asset j's returns in week t_w and the existence of granularity in social

attention in that week (as defined above). The effect indicates that granularity in social discussions are linked to a 0.8pp increase in asset volatility in a given week, on average.

Table 6.3: Relationship between social granularity and volatility

	Dependent variable: σ^2_{j,t_w}
Granularity Indicator	0.0076 *(0.000)* ***
σ^2_{j,t_w-1}	0.126 *(0.001)* ***
r_{j,t_w}	0.000 *(0.000)*
Week FE	Yes
Number of obs.	2,479,664
Adjusted R^2	0.029

Notes: the dependent variable is the variance in the log returns of asset j in week t_w, σ^2_{j,t_w}. Explanatory variables include: the variance in log returns in the previous week σ^2_{j,t_w-1}, the average log returns in the present week, r_{j,t_w}, as well as an indicator for whether we observe social granularity in the week (the indicator equals one if the ticker receives five posts in a given week and a pareto-fit indicates that the distribution of popularity is heavy-tailed). The sample includes all tickers discussed on WSB since 2016; the sample period begins in 2016 and ends at our WSB cutoff time. Accompanying standard errors, displayed in brackets, are clustered at the stock level.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

The analysis is only a preliminary indication and does not imply causality. Furthermore, we cannot perform the same test for returns, since the direction of idiosyncratic demand (long or short) is important to quantify. Identifying the link between granular social shocks and returns is tricky due to confounding variables – returns and popularity of online content about an asset may both be driven by news or other market factors. The next step of our empirical approach therefore consists of extracting idiosyncratic social shocks, measured as unexplained idiosyncratic variation in the sentiments of submissions.

Estimates of idiosyncratic social shocks To extract unexplained variation in sentiments, we regress the sentiment expressed in a given post on the return on day t, $r_{j,t}$, the cumulative returns on the week containing day t, r_{j,t_w}, and average sentiments expressed by peers in the prior week $\bar{\Phi}_{j,t_w-1}$. Since our analysis includes weekly and daily variables to better extract idiosyncratic social shocks, we distinguish between t time in days, and t_w time in weeks. For

a post about asset j made by author i at time t, we estimate the idiosyncratic information content of the post as $e_{i,j,t}$ in the following regression:

$$\Phi_{i,j,t} = \beta_1^A r_{j,t} + \beta_2^A r_{j,t_w} + \beta_3^A X_{t_w} + \beta_4^A \bar{\Phi}_{j,t_w-1} + e_{i,j,t}, \tag{6.3}$$

where X_j are asset fixed effects. These control for the persistence in sentiments for certain assets – for example, the forum's enthusiasm about TSLA. We note that the results remain similar with and without ticker fixed effects.

The strategy follows the reasoning outlined in Section 4.2.1, Figure 4.B.1. We posit that any news that emerges at time t about a company is assimilated by the market and manifests in returns. Any variation in sentiment that is unexplained by market performance at time t and by past discussions is submission-specific and is idiosyncratic to news and information more broadly available about that stock at that time.

The object of interest, residual $e_{i,j,t}$, is information shared in the submissions that is orthogonal to asset j's returns at time t or within week t_w. We, therefore, would expect $e_{i,j,t}$ to have an impact on the market through social forces, rather than through purely informational content. Figure 6.2 plots the distribution of idiosyncratic social shocks. The distribution is somewhat asymmetric and the modal, unexplained sentiment is bullish, but the left tail of discussions demonstrates the presence of intense bearish discourse.

Estimating the effect of granular attention In order to assess the impact on asset returns of granular social attention, we proceed by analyzing the following relationship:

$$r_{j,t_w+1} = \beta_1^B \bar{e}_{j,t_w} + \beta_2^B r_{j,t_w} + \beta_3^B X_{t_w} + v_{j,t_w}, \tag{6.4}$$

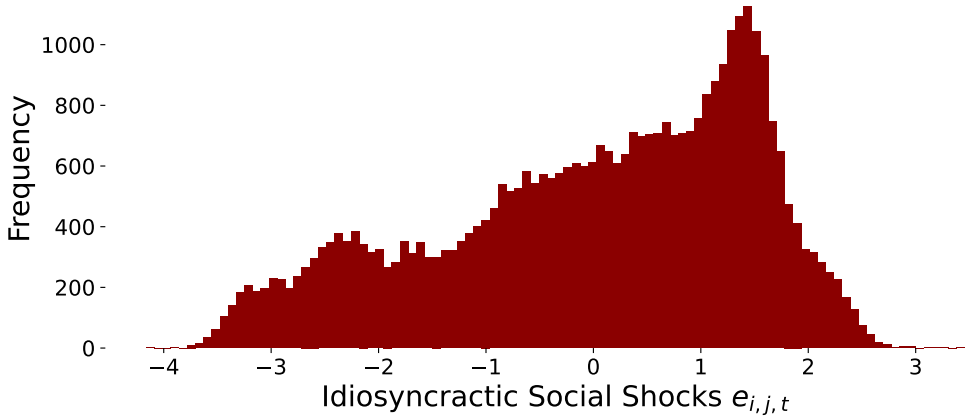


Figure 6.2: **Idiosyncratic social shocks:** this graph plots the distribution of idiosyncratic sentiment heterogeneity, $e_{i,j,t}$, from Eq. 6.3.

where X_{t_w} are week fixed effects, r_{j,t_w+1} is the cumulative return of stock j in week t_w+1, r_{j,t_w} is the cumulative return in asset j in week t_w, $\bar{e}_{j,t_w} = \sum_i s^c_{i,j,t_w} e_{i,j,t_w}$ the popularity-weighted average idiosyncratic sentiment shared about stock j in week t_w ($\sum_i s^c_{i,j,t_w} = 1$), and v_{j,t_w} is a stock-week specific error. s^c_{i,j,t_w} is calculated by summing the total number of comments received across all posts in stock j in week t_w – the comment count for a specific submission is then normalized by the total comment count within week t_w. Comment count is re-indexed so that the minimum number of comments a post receives is one. We choose to model future returns in order to mitigate any confounding variables.

A key identification challenge stems from our goal to identify the impact of social forces and popularity of some content over other content, versus general idiosyncratic sentiment on the forum. We rely on a GIV for our identification strategy. The GIV is defined as the difference between popularity-weighted and equally-weighted social shocks, each aggregated for the stock j at period t_w:

$$GIV_{j,t_w} = \sum_i s^c_{i,j,t_w} e_{i,j,t_w} - \sum_i \frac{1}{N_{j,t_w}} e_{i,j,t_w}, \tag{6.5}$$

where N_{j,t_w} is the total number of posts about stock j in week t_w, and i at the authors who are active in the discussion about asset j in week t_w. We subsequently replace $\bar{e}_{j,t_w}$ in Eq. 6.4

with $\hat{u}_{j,t_w}$, where $\hat{u}_{j,t_w}$ is the predicted values from the regression of the GIV on the social shocks $\bar{e}_{j,t_w}$. The outcome variable $\hat{u}_{j,t_w}$ is driven by the popularity of certain posts over others, rather than by general idiosyncratic sentiment.

GIV requirements and threats to identification In addition to capturing the impact of social forces, our specification allows us to disentangle asset properties which affect both sentiment and returns simultaneously. The use of the GIV allows us to mitigate the common shocks problem, where certain stock-specific shocks could affect all idiosyncratic sentiments and future returns – this is similar to Galaasen et al. (2020) in reasoning. Specifically, stock j's returns in week $t_w + 1$ may be driven in-part by social forces, but also by asset properties which affect both sentiment and returns, which we are unable to control for while estimating idiosyncratic sentiments in Eq. 6.3. A correlation of these shocks with $\bar{e}_{i,t_w}$ may result for a biased estimator for β_1^B. More formally, outcome variable, after imposing controls from Eq. 6.4, y^r_{j,t_w+1} (return variable in week t_{w+1} renamed to demonstrate the incorporation of controls), may be of the form:

$$y^r_{j,t_w+1} = \beta_1^B \bar{e}_{j,t_w} + \eta_{j,t_w},$$

where η_{j,t_w} is the 'common shock' to asset j in period t_w.

We assume that the idiosyncratic social shock from a post can be expressed as having a stock level component, common to all posts about the stock within that time period, and a post-level component:

$$e_{i,j,t_w} = \beta_j^{cs} \eta_{j,t_w} + u_{i,j,t},$$

where β^{cs}_{j,t_w} is the sensitivity of posts within week t_w to the common shock to stock j.

The identification strategy rests on the assumption that the popularity of the idiosyncratic content of posts $s^c_{i,j,t_w} e_{i,j,t_w}$ is not correlated with stock-week shocks η_{j,t_w}. More formally, we require $E[u_{i,j,t}\eta_{j,t_w}] = 0$: the idea is that there are social shocks which make certain content on WSB popular over other content, but that is orthogonal to shocks affecting asset j in week t_w. This is not a problem in this setting for several reasons. Firstly, the popularity of a post could potentially be linked to a stock through it's informativeness about the asset's price. However, in the creation of our social idiosyncratic shocks, we extract shocks while controlling for returns on the day and the week of the post. Our social shock time series is, therefore, orthogonal to asset returns at time t and in week t_w and is, therefore, orthogonal to new information available to investors at the time. Furthermore, we find post popularity to be uncorrelated with asset returns on day t and week t_w on which the post is made. As a final precaution, we look at the time period during which a post is active on WSB, where the final time that a post is active is the final comment activity on the post (if the post receives no comments, it is simply the time of the post). We remove posts from our sample that receive commenting activity into the week following the post. On WSB we observe data about relatively unsophisticated retail investors where the sentiments of posts at time t about an asset are systematically linked to *negative* future returns (this holds both when we take a raw average and popularity-weighted average average sentiment). This is additional proof that investors we observe do not have access to information on stock-level shocks. Finally, both our shock and popularity time series are constructed at time t_w, while the dependent variable is observed at time $t_w + 1$, avoiding contemporaneity issues.

Results In order to study the impact of granular social attention, we run the following regression on weekly returns and posts:

$$r_{j,t_w+1} = \tilde{\beta}^B_1 \hat{u}_{j,t_w} + \beta^B_2 r_{j,t_w} + \beta^B_3 X_j + v_{j,t_w},$$

Table 6.4: Stock returns versus granular social shocks

	Dependent variable: r_{j,t_w+1}				
	Average $\bar{e}_{j,t_w}$	Popularity-weighted $\bar{e}_{j,t_w}$	Instrumented by GIV: $\hat{u}_{j,t_w}$		
	(1)	(2)	(3)	(4)	(5)
Granular Social Shock	0.007	0.010***	0.012***	0.013***	0.011**
	(0.005)	*(0.004)*	*(0.005)*	*(0.005)*	*(0.005)*
r_{j,t_w}	-0.099***	-0.100***	-0.059***	-0.102***	-0.052*
	(0.037)	*(0.037)*	*(0.012)*	*(0.037)*	*(0.028)*
Ticker FE	Yes	Yes	No	Yes	No
Week FE	No	No	No	No	Yes
Observations	2,201	2,201	2,201	2,201	2,201
R^2_{adj}	0.160	0.164	0.013	0.163	0.124

Notes: this table presents the OLS estimates for the influence of idiosyncratic social shocks on WSB. Columns (1) and (2) present the effect of average idiosyncratic shocks and popularity-weighted idiosyncratic shocks, respectively. Columns 3-5 present various specifications of our instrumented idiosyncratic social shocks $\hat{u}_{j,t_w}$. Robust standard errors, clustered at the ticker level, are shown in parentheses. The F-statistic for the first stage regression is 2,297. Observations with incomplete data are dropped.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

where $\hat{u}_{j,t_w}$ are the *fitted values* of idiosyncratic social shocks on our GIV (we rename our coefficient to $\tilde{\beta}_1^B$ to highlight this), r_{j,t_w+1} is the cumulative weekly return of stock j in week t_w+1, X_j are ticker fixed effects. The formulation closely follows that of our extended model with granular shocks, from Eq. 5.5.

Several patterns emerge from our empirical exercise, presented in Table 6.4. Firstly, we observe that the data appears to follow the structure proposed in our model – idiosyncratic social shocks are positively linked to future returns. In column (4), the estimated effect can be summarised as follows: an estimate for $\tilde{\beta}_1^B$ at 0.013, means that the idiosyncratic doubling in the odds of a very popular post expressing bullish over bearish sentiments (while less popular posts do not express an idiosyncratic sentiment) increases returns in the following week by one percent, on average. The effect is small, but persists across specifications. Consistently with our model prediction, the average idiosyncratic noise, in column (1), has no effect.

We explore the data in greater detail in Appendix 6.A. We consider instances where we observe a large (in magnitude) $\hat{u}_{j,t_w}$ and the types of posts that triggered the viral discussions. We observe that the viral, influential content has very different style. Some posts present a thorough, substantiated discussion backing the author's claim. Others attract attention through the large risk the author has taken, potentially inspiring others to follow in their footsteps.

6.4 Evidence of bubbles

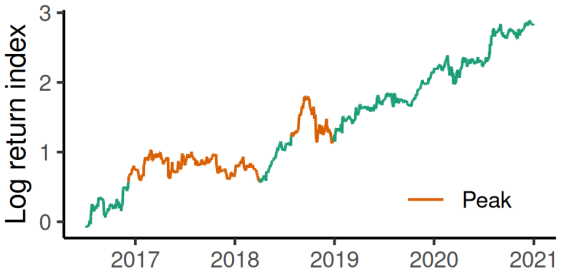
Our modeling (section 5.2) exercise demonstrates that, in the presence of peer effects, bubbles can arise due to purely social forces. In this section, we ask whether we can identify Reddit's bull runs. Greenwood et al. (2019) propose a transparent classification scheme to identify potential bubbles. They determine time windows in which the price indices for various industry market capitalisations grew at a 'rapid' rate, constituting a sample of 'run-ups'. These run-ups are then separated into those whose price levels remained constant, versus those whose price levels crashed (where the rate of increase and price level of the crash are pre-selected). A run-up followed by a crash constitutes an instance in which the index experienced a 'bubble'.

We adapt the method of Greenwood et al. (2019) to identify large run-ups in stock prices, and test whether activity on Reddit during the price run-up is related to an eventual downturn. The two main variables we use to measure social activity are the number of submissions that mention a given stock in the month preceding a price run-up, as well as the average sentiment expressed in that time window.

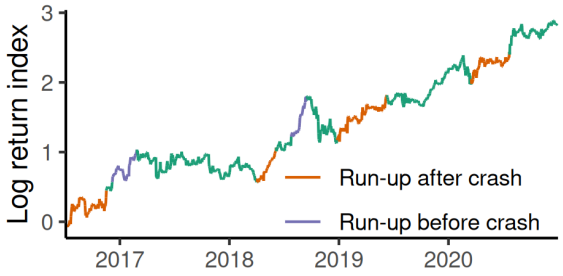
The challenge to finding price run-ups for *individual stocks* is that large price swings are more erratic compared to the returns on broad industries considered by Greenwood et al. (2019). Importantly, run-ups in stock prices can be immediate under small market

capitalisation, but slower with large market capitalisation. An arbitrary condition on time and return magnitude thus introduces considerable selection concerns into a sample of run-ups in stock prices. This calls for a more flexible method.

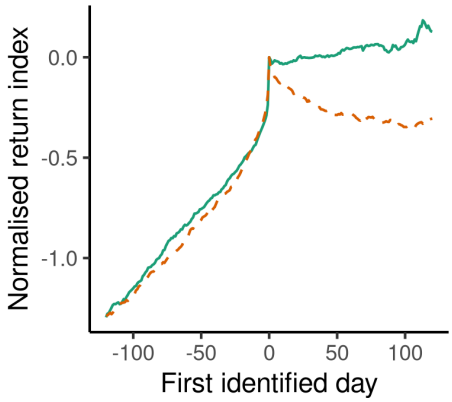
We tweak the methodology from Greenwood et al. (2019), but still rely on their theoretical framework: the identification of price run-ups, in relation to corresponding price crashes. The goal of our method is specifically to find instances when a price run-up either precedes or follows a subsequent price decline.



(a) AMD Peaks in returns



(b) AMD Run-ups relative to crashes



(c) Run-up dynamics

Figure 6.3: **Price run-ups identification**; Figures 6.3a, 6.3b display identified peaks and run-ups in AMD. Figure 6.3c displays run-up dynamics averaged across our sample: returns following a 'recovery' run-up, in solid green, remain stable, while they substantially decline after a run-up followed by a crash.

Identifying price run-ups In order to remain as systematic as possible, we identify 'excursions' from local minima in the time series of each stock's log cumulative return index. We define excursions as observations that precede a future minimum, starting from a previous point that matches the cumulative return at that minimum. Specifically, this future

minimum is a price level to which the stock does not return at any point following that date. Intuitively, the price of those stocks reverts to a future minimum at the start of the excursion, and thus constitutes a 'peak'. Here, we filter excursions with a maximum return from the minimum of over 50% as peaks. Figure 6.3a includes two examples of peaks for AMD, one slow run-up starting in late 2016, and an abrupt one in mid 2018.

In addition to peaks, we can also define troughs as excursions from local maxima. An excursion from a local maximum, as opposed to a minimum, is an observation with a cumulative return lower than the highest value in the stock's history. As such, these function as 'inverted' peaks: a series of negative returns following a high price, which ends when the local maximum is recovered. As with peaks, we isolate troughs with a maximum cumulative return over 50% from the lowest return to the following local maximum.

In summary, both peaks and troughs constitute price swings of over 50% in either direction, one with a period of negative returns (troughs) and one of positive returns (peaks). Peaks and troughs differ in the ordering of price run-ups versus crashes: for peaks, the run-up is followed by a crash, whereas the crash is followed by a run-up recovery period for troughs. In this fashion, we can leverage Greenwood et al.'s (2019) classification of bubbles by studying run-ups that precede crashes (during a peak) and run-ups that follow a crash (during a trough). We illustrate this distinction for the AMD price series in Figure 6.3b, where segments in purple correspond to the run-up periods during bubbles (where peaks are highlighted in orange Figure 6.3a), whereas segments in orange in Figure 6.3b represent run-ups that follow crashes.

We implement this procedure for all stocks traded on the NYSE, NYSE Mkt and NASDAQ, with share codes 10 and 11. In what follows, we restrict the sample of peaks to those after January 1, 2016, and for stocks with an average market capitalisation over one billion USD during the peak. Finally, to harmonise all instances of run-ups with varying lengths

of time, we filter out run-ups with an average daily annualised return less than 200% (corresponding to 0.29% daily) to match the criterion used by Greenwood et al. (2019). This leaves us with a sample of 329 run-ups, of which 31 occur during a 'peak' period and precede a crash. We visualise the average cumulative return index by type of run-up in Figure 6.3c. Importantly, the trends prior to the turning point are similar for run-ups followed by downturns, and run-ups following downturns.

Table 6.5: WSB features for price runups and crashes

	Run-up after downturn (1)	Run-up before crash (2)	**Difference** (3)	**T test** (4)
Log Mentions	0.02 *(0.19)*	0.19 *(0.63)*	0.17	15.27
Sentiment	0.02 *(0.31)*	0.13 *(0.52)*	0.10	11.41
No. Observations	39,392	3,410		

Notes: this table presents the prevalence of WSB discussions in assets that experience bubble-like dynamics - the mean for each stock sample is shown, with its standard deviation in parentheses. The first column displays that discussions are limited in stocks that are in recovery and do not experience a downturn; however, both mentions and sentiment are higher for stocks that experience a run-up and a subsequent crash, as shown in column (2). A single observation is a day during a run-up period.

Results For each type of run-up – before or after a crash – we count the number of times a submission is made on WSB in the corresponding time window, mentioning the stock experiencing the run-up. Similarly, we compute the average sentiment expressed in these submissions. Table 6.5 summarises our findings, along with the corresponding standard deviation. The striking result is that WSB activity, in both the log of mentions as well as the average sentiment, features significantly more prominently in run-ups before crashes, rather than after crashes – the difference in means is highly significant.

Are users similarly aware of both types of stocks? One might argue that stocks that experience a crash may have some underlying characteristics which make them of greater interest to the WSB crowd. The share of run-ups preceding crashes during which the corresponding stock is mentioned on WSB, 64.52%, is close to the same share for run-ups follow-

ing crashes, at 54.36%. The samples are thus similarly represented in terms of discussions in WSB. Figure 6.3c demonstrates that the average price trends for both types of run-ups are nearly identical before the cutoff date, then subsequently diverge.

Summary Our findings demonstrate a link between bubble-like dynamics in the markets and discussions among investors, and prompt additional important questions for research. Recently, rich datasets have become available detailing individual investor portfolios, allowing the study of investor attributes and portfolio choices (Balasubramaniam et al. 2023), individual extrapolation during asset price bubbles (Pearson et al. 2021), and other important characteristics. Combining discussion data and portfolio data offers a promising venue for further research into the profit and loss profiles of hype investors, as well as the behaviour of other market participants faced with hype investor demand.

Appendix

6.A GIV

Extracting idiosyncratic social shocks In Table 6.A.1, we present the coefficients from Eq. 6.5. We observe that, consistently with our previous findings and the proposed structure in Figure 4.B.1, returns and past user sentiments all impact the expressed sentiment of user i about asset j at time t.

Table 6.A.1: Idiosyncratic social shocks estimation

	Dependent variable: $\Phi_{i,j,t}$
$r_{j,t}$	0.324 (0.086) ***
r_{j,t_w}	0.370 (0.090) ***
Φ_{j,t_w}	0.212 (0.043) ***
Ticker FE	Yes
Number of obs.	46,694
R^2	0.077
F-statistic	41.67

Notes: The dependent variable is the log-odds of a given submission by author i at time t on stock j to express bullish over bearish sentiment. Explanatory variables include: the log return on day t, $r_{j,t}$, the log-return in week including day t denoted as t_w, r_{j,t_w}, and the average past sentiment of peers, $\bar{\Phi}_{i,j,t_w-1}$. Accompanying standard errors, displayed in brackets, are clustered at the stock level.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

Table 6.A.2: GIV - First Stage

	Dependent variable: $\bar{e}_{j,t_w}$
GIV	0.932 (0.019) ***
Number of obs.	2,441
R^2	0.510
F-statistic	2,292

Notes: The dependent variable is the popularity-weighted average idiosyncratic sentiment expressed about asset j in week t_w, $\bar{e}_{j,t_w}$. The explanatory variable is the difference between the popularity-weighted and raw average of idiosyncratic sentiments expressed about asset j in week t_w.
*** Significant at 1% level ** Significant at 5% level * Significant at 10% level

GIV - First Stage In Table 6.A.2, we present our First Stage regression, where we regress the popularity-weighted idiosyncratic sentiment, as our dependent variable, on the differ-

ence between the popularity-weighted and regular average idiosyncratic sentiment. We observe that our GIV is highly predictive of the popularity-weighted sentiment, $\bar{e}_{j,t_w}$, however, explains only part of the variation. In this way, we are able to extract the element of the popularity-weighted sentiment measure driven by social preferences, versus those driven by other factors.

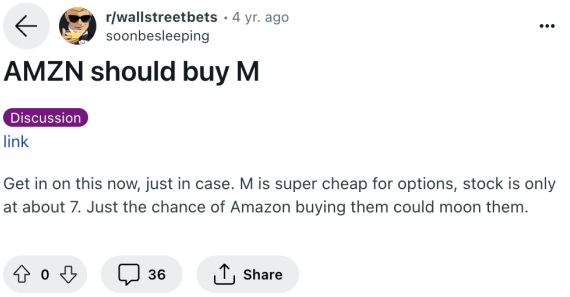
We illustrate the effect of the GIV through the following two scenarios. In *Scenario One*, a very popular and an unpopular post both express a positive, idiosyncratic sentiment. In *Scenario Two*, a very popular expresses the positive sentiment, while the unpopular posts expresses no idiosyncratic sentiment. Without employing the GIV, our original popularity-weighted average idiosyncratic sentiment measure $\bar{e}_{j,t_w}$ would be very similar in both scenarios. However, by using the GIV, we would *predict* zero idiosyncratic sentiment in *Scenario One*, but a large positive idiosyncratic sentiment in *Scenario Two*. In this way, our GIV allows us to distinguish the signal coming from popularity.

Influential posts – examples We present some examples of popular content during weeks with a with a large $\hat{u}_{j,t_w}$. Figure 6.A.1 presents the most popular posts about AMZN in the week of June 24-26, 2020 – the posts were identified as having large, positive, idiosyncratic sentiment within the week. AMZN exhibited a return of 7.8% the following week. Figure 6.A.3 presents the most popular posts about WYNN in the week of March 16-20, 2020 – the post had a positive, idiosyncratic sentiment within the week. WYNN exhibited a return of 21.8% the following week. Figure 6.A.2 presents the most popular post about GM in the week of March 9-13, 2020 – the post was identified as having large, negative, idiosyncratic sentiment within the week. The following week, GM exhibited a 27% decline.

We observe that the content has very different style. Some discussions, such as the GM post in Figure 6.A.2, present a thorough, substantiated discussion backing the author's claim. Other popular content, such as the WYNN post in Figure 6.A.3, attracts attention

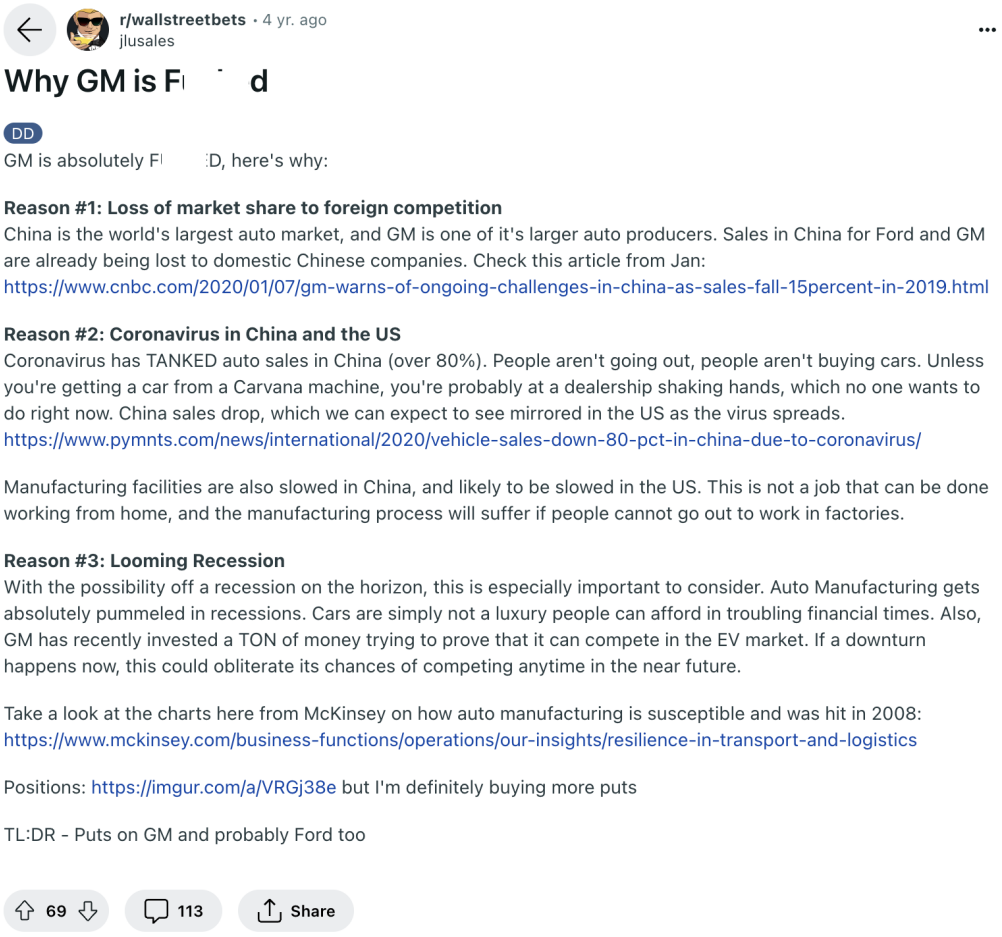
through the large risk that the author has taken, potentially inspiring others to follow in their footsteps.

(a) Popular post about AMZN: June 20, 2020 (redacted to remove inappropriate content)



(b) Popular post about AMZN: June 24, 2020

Figure 6.A.1: **Popular post about AMZN: June, 2020**



r/wallstreetbets • 4 yr. ago
jlusales

Why GM is F d

DD

GM is absolutely F :D, here's why:

Reason #1: Loss of market share to foreign competition
China is the world's largest auto market, and GM is one of it's larger auto producers. Sales in China for Ford and GM are already being lost to domestic Chinese companies. Check this article from Jan:
https://www.cnbc.com/2020/01/07/gm-warns-of-ongoing-challenges-in-china-as-sales-fall-15percent-in-2019.html

Reason #2: Coronavirus in China and the US
Coronavirus has TANKED auto sales in China (over 80%). People aren't going out, people aren't buying cars. Unless you're getting a car from a Carvana machine, you're probably at a dealership shaking hands, which no one wants to do right now. China sales drop, which we can expect to see mirrored in the US as the virus spreads.
https://www.pymnts.com/news/international/2020/vehicle-sales-down-80-pct-in-china-due-to-coronavirus/

Manufacturing facilities are also slowed in China, and likely to be slowed in the US. This is not a job that can be done working from home, and the manufacturing process will suffer if people cannot go out to work in factories.

Reason #3: Looming Recession
With the possibility off a recession on the horizon, this is especially important to consider. Auto Manufacturing gets absolutely pummeled in recessions. Cars are simply not a luxury people can afford in troubling financial times. Also, GM has recently invested a TON of money trying to prove that it can compete in the EV market. If a downturn happens now, this could obliterate its chances of competing anytime in the near future.

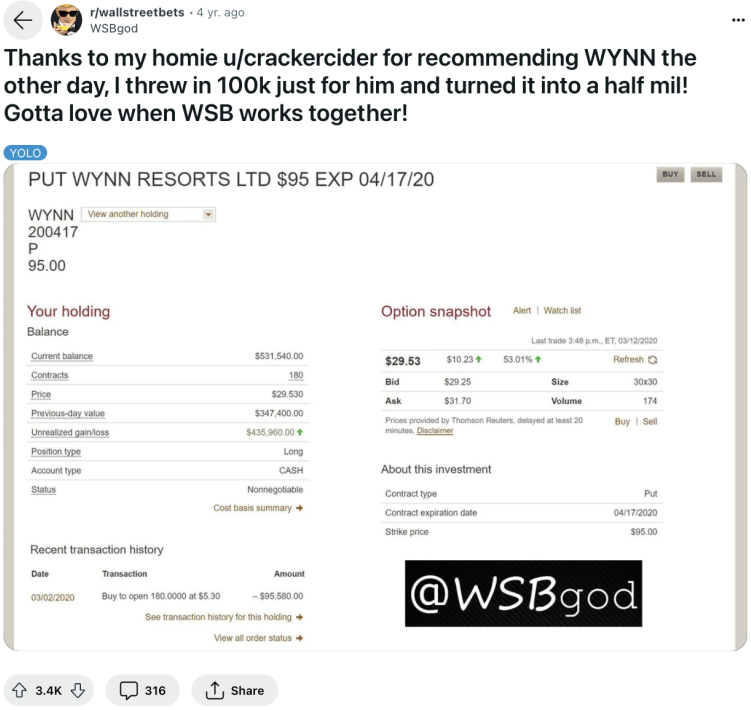
Take a look at the charts here from McKinsey on how auto manufacturing is susceptible and was hit in 2008:
https://www.mckinsey.com/business-functions/operations/our-insights/resilience-in-transport-and-logistics

Positions: https://imgur.com/a/VRGj38e but I'm definitely buying more puts

TL:DR - Puts on GM and probably Ford too

69 113 Share

Figure 6.A.2: **Popular post GM: March 10, 2020 (redacted to remove inappropriate content)**



r/wallstreetbets • 4 yr. ago
WSBgod

Thanks to my homie u/crackercider for recommending WYNN the other day, I threw in 100k just for him and turned it into a half mil! Gotta love when WSB works together!

YOLO

PUT WYNN RESORTS LTD $95 EXP 04/17/20

BUY SELL

WYNN View another holding
200417
P
95.00

Your holding
Balance

Current balance	$531,540.00
Contracts	180
Price	$29.530
Previous-day value	$347,400.00
Unrealized gain/loss	$435,960.00
Position type	Long
Account type	CASH
Status	Nonnegotiable

Cost basis summary

Recent transaction history

Date	Transaction	Amount
03/02/2020	Buy to open 180.0000 at $5.30	– $95,580.00

See transaction history for this holding
View all order status

Option snapshot Alert | Watch list

Last trade 3:48 p.m., ET, 03/12/2020

$29.53	$10.23	53.01%	Refresh
Bid	$29.25	Size	30x30
Ask	$31.70	Volume	174

Prices provided by Thomson Reuters, delayed at least 20 minutes. Disclaimer

Buy | Sell

About this investment

Contract type	Put
Contract expiration date	04/17/2020
Strike price	$95.00

@WSBgod

3.4K 316 Share

Figure 6.A.3: **Popular post WYNN: March 13, 2020**

Chapter 7: Social Interactions and Temporal Network Dynamics

The previous chapters establish the importance of the dynamics on social media within the financial context. In this chapter, we expand our analysis by looking at how several political and social movements have evolved over time. Due to the lack of high-quality, annotated data tracking online conversations (and the challenge posed to LLMs in this task), the first part of the chapter presents a new dataset tracking agreements and disagreements between authors across several Reddit forums. The second part presents a novel, signed temporal clustering algorithm and applies it to social media data and international news data, with the goal of understanding how online discussions and international communities have evolved over time.

7.1 DEBAGREEMENT: a dataset of agreements and disagreements in online political discussions

Online debates have a considerable impact on society. With over 4.2 billion people actively using social media, gaining insights into the evolution of online discussions and movements is important for explaining social change.[1] Fortunately, the field of NLP offers techniques to understand textual interactions. Specifically, one important research area is *(dis)agreement detection*, as it is fundamental for understanding societal polarisation and the spread of ideas online (Ribeiro et al. 2017, Wojcieszak 2011, Tan et al. 2016, Rosenthal & McKeown 2015).

(Dis)agreement detection falls under the field of *stance detection* (Walker et al. 2012) – the automatic classification of the position (or stance) of the producer of a piece of text, towards a target, into one of three classes: *in favor*, *against*, or *neutral*. Due to the explosion of available online data sources, there has been a plethora of automatic natural language

[1] https://datareportal.com/reports/digital-2021-global-overview-report

systems aimed at detecting stances. They have used either feature-based machine learning, deep learning, or ensemble learning approaches. A comprehensive review of these systems is presented in (Küçük & Can 2020), section 5. One noticeable trend is the adaptation of pre-trained language representation models for stance detection, such as BERT (Devlin et al. 2018), RoBERTa (Liu et al. 2019), DeBERTa (He et al. 2020) and XLNet (Yang et al. 2019), since they have led to considerable performance improvements for NLP tasks.

Although researchers have initially modelled stances using only text, recent work has shown that stance detection would benefit from context-sensitive approaches. In particular, several methods have leveraged graph (or network) features, such as interaction networks, preference networks, and connection networks. Darwish et al. (2018, 2020), Borge-Holthoefer et al. (2015) use retweet data, while Samih & Darwish (2021), Dey et al. (2017) leverage hashtags to infer Twitter users' stances. Researchers have also explored how to incorporate social context (Keskar et al. 2019, Kulkarni et al. 2021) and structured knowledge (Colon-Hernandez et al. 2021) into language models to improve inference on NLP tasks.

Despite the growing literature leveraging contextual features for stance detection, most existing datasets provide only textual information. The few datasets which provide contextual information are tweet datasets. These suffer from several drawbacks: i) they are shared via tweet identifiers, making it impossible to retrieve deleted tweet content and network information of deleted users for future research, and ii) retweets and hashtags may ease stance detection, but are features specific to Twitter discussions.

Contributions We introduce DEBAGREEMENT, a dataset for detecting (dis)agreements in real-world online discussions. The dataset contains 42,894 comment-reply pairs, as well as contextual information (authorship, post, timestamp, etc), extracted from `reddit`. This dataset presents opportunities to detect (dis)agreements by leveraging context beyond text

and does not rely on platform-specific features, such as retweets or hashtags. Unlike existing datasets for stance detection, DEBAGREEMENT provides realistic online discussions, with diverse writing styles, genres and topics of discussion. We evaluate state-of-the-art (SOTA) pre-trained Language Models (LMs) on DEBAGREEMENT: our findings highlight ways to improve LMs with contextual information, and emphasize the substantial difference between DEBAGREEMENT and existing datasets. DEBAGREEMENT is available to download at: https://scale.com/open-datasets/oxford. Details of what data is contained within the dataset are available in the supplementary material.

Impact DEBAGREEMENT presents new opportunities for modeling diverse online interactions with text and context (authorship, graph, temporal information). Reddit's popularity, and the fact that all reddit discussions are downloadable for research purposes, make this a valuable data source to invest resources into understanding. Furthermore, the graph structure provided by DEBAGREEMENT offers opportunities for combining text-based machine learning (ML) and graph representation learning (GRL) methods. Modeling online discussion forums as graphs of interactions between users enables researchers to: i) translate the (dis)agreement detection task into a sign link prediction one (Zhang et al. 2019), ii) use recent advances in GRL methods (Hamilton 2017), and iii) leverage existing signed graph embedding methods (Wang, Tang, Aggarwal, Chang & Liu 2017, Derr et al. 2018, Rahaman & Hosein 2018) tested on publicly available signed graphs, such as Epinions and Slashdot (Leskovec et al. 2010*b*,*a*, Tang et al. 2016).

DEBAGREEMENT follows active reddit users over time across five forums: r/BlackLivesMatter, r/Brexit, r/climate, r/democrats, r/Republican. It provides the possibility to investigate social theories, understand polarisation, and study how people express their opinions and change their views on social media.

7.1.1 Related datasets

Küçük & Can (2020) present a survey of recent advancements in stance detection and discuss 24 publicly available text datasets for stance detection (summarized in Table 6 of the survey). Among them, only two have more than 7,000 annotations (Darwish et al. 2017, Dean Pomerleau 2017) and two are annotated for (dis)agreement detection specifically (Baly et al. 2018, Dean Pomerleau 2017). The currently available datasets, summarised in Küçük & Can (2020), have several drawbacks: i) they only provide textual data, ii) all but two have fewer than 7,000 annotations, iii) they span a small range of topics and text genres, and iv) their genre and writing style of the text is often formal and structured. The latter point often simplifies the task of stance detection, however, is not representative of online debates on popular platforms. For instance, in the popular Perspectrum dataset (Chen et al. 2019), claims (e.g. *Animals should not be used for scientific or commercial testing*) and perspectives (e.g. *Any living entity should not be treated as objects or property as doing so allows them to be treated amorally*) are presented in a formal writing style.

In the literature, only two publicly available datasets contain contextual data along with the text and are annotated for stance detection in social media . Taulé et al. (2018) provide 11,398 annotated Spanish tweets related to the 2017 Catalan Referendum (in favor of the independence, against, or neither), as well the images and tweets before and after these tweets on users' timelines. Cignarella et al. (2020) present 3,282 annotated Italian tweets related to the 2019 sardines movement. The authors provide contextual information based on the tweet (number of retweets, likes, replies and quotes received to the tweet, type of posting device, date), the tweet's author (number of followers, of tweets ever posted, user's bio) and their social network (friends, replies, retweets, quotes' relations). However, as pointed out by Küçük & Can (2020), AlDayel & Magdy (2021), these datasets suffer from the following drawbacks: i) tweet datasets are shared via tweet identifiers, making it impossible to re-

trieve deleted tweet content and network information of deleted users for future research, and ii) retweets and hashtags may ease stance detection, but are features specific to Twitter discussions. Furthermore, Twitter does not provide complete, open-source data access for research purposes.

7.1.2 Debagreement

Collection

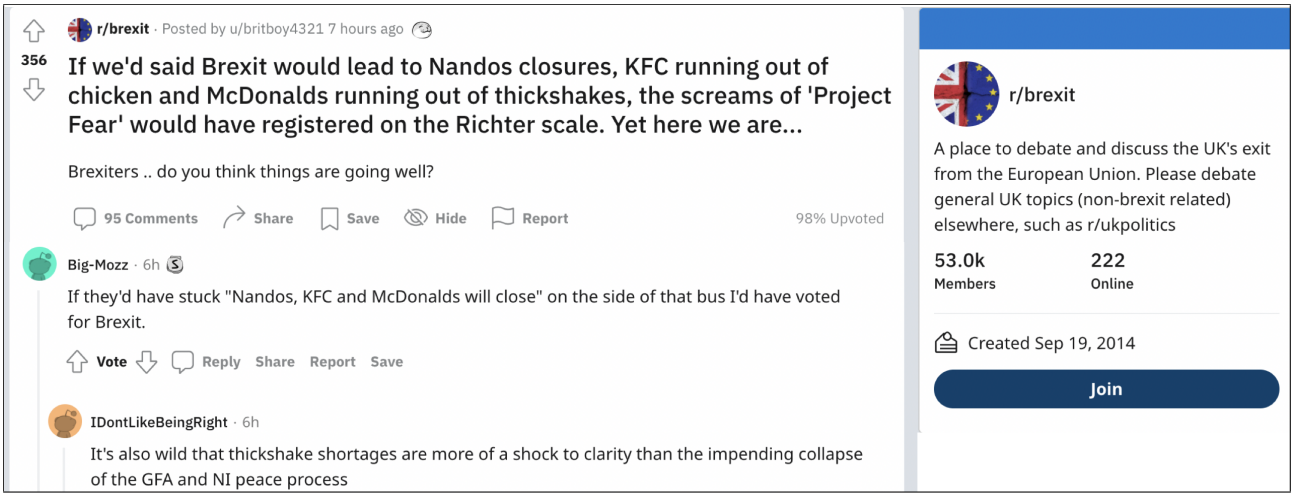


Figure 7.1: r/Brexit

DEBAGREEMENT was created by retrieving data from a popular discussion forum website, reddit. Reddit is broken down into discussion sub-forums, called subreddits, and denoted r/* (e.g. r/Brexit). Within each subreddit, users write titled posts, typically accompanied with a body of text and/or a link to an external website. These posts (or submissions) can be commented, *upvoted* or *downvoted* by other users (Figure 7.1). A ranking algorithm raises the visibility of a submission based on the number of upvotes it receives, but lowers it with time, so that the first posts visitors see are highly upvoted and/or recent. Comments related to a post are also visible, and are subject to a similar scoring and commenting system.

We collected reddit data from the PushShift API.[2] The API provides historical reddit data, which is open-source for researchers (Baumgartner et al. 2020). We pull data from

[2] https://pushshift.io/

three subreddits of social movements r/Brexit, r/BlackLivesMatter, r/climate, and two subreddits of political affiliations r/Republican, r/democrats. A description of each subreddit, summarised from the subreddit websites, along with usage counts (as of August 2021), are presented below:

- r/BlackLivesMatter discusses news related to the *Black Lives Matter* movement. It was created in 2014 and has 109K members,
- r/Brexit aims to foster debate about the United Kingdom's (UK) exit from the European Union (EU). It was created in 2014 and has 53K members,
- r/climate is a community for truthful science-based news about climate and related politics and activism. It was created in 2008 and has 99K members,
- r/democrats is a partisan subreddit. It aims to discuss political news, policies and how to ensure the election of Democratic party candidates. It was created in 2014 and has 292K members,
- r/Republican is a partisan subreddit for Republicans to discuss issues with each other. It was created in 2008 and has 172K members.

For r/climate and r/Brexit, we collect all the submissions and posts since the subreddits' creation dates until May 2021, in order to track these social movements from their inception. For r/BlackLivesMatter, r/democrats and r/Republican, we collect data from January 2020 in order to focus on recent, critical events: the protests following George Floyd's murder and the 2020 United States presidential election.

Data cleaning We first excluded empty comments, comments from deleted authors, comments that were hidden for user privacy reasons, and comments containing hyperlinks. Inside the text body of both the submissions and the comments, we removed paragraph breaks

and replace a small set of special characters (for example, *&amp* is replaced with *and*). Despite potentially offensive content being present in the discussion forums, we purposefully do not filter out this content in order to realistically capture the nature of online discussions.

High-quality interactions A large amount of submissions receive no or few comments on reddit. Respectively 92%, 45%, 83%, 75%, and 59% of posts received less than 5 comments in r/BlackLivesMatter, r/Brexit, r/climate, r/democrats, and r/Republican. In order to annotate impactful discussions in a given subreddit, we remove posts with fewer than 10 words and with fewer than k comments, where k is the rounded average number of comments per submission on each forum ($k = 2, 5, 5, 5, 10$ for r/BlackLivesMatter, r/climate, r/Republican, r/democrats, and r/Brexit, respectively). We also filtered out comments with fewer than 10 or more than 100 words, and comments that contain hyperlinks. We truncate submissions to 100 words in length, as this is sufficient to contextualize the interaction for annotators. After further inspecting the nested discussion structure on reddit, it became apparent that discussion threads between users with different opinions often devolve into an onslaught of negative affronts with little substance. Conversely, conversations between people with similar opinions tend to contain only a few nested comments. These observations motivated our decision to retain only comment-reply pairs whose parent comment replied directly to the initial submission (nest level 1), as proxies of lengthier discussion threads between two given users. We also remove comments whose authors have been deleted or whose contents have been removed. The percentage of data affected by each filtering step is provided in the supplementary material.

Graph creation For each subreddit r/$*$, the resulting set of interactions forms a multi-edge, temporal graph $\mathcal{G}_{r/*}$, where nodes are users, and edges represent a comment-reply interaction between two users. One of the unique advantages of DEBAGREEMENT over other

datasets is the additional graph interaction information provided about every subreddit.

For all of the forums except r/Brexit, we keep all comment-reply pairs as the final dataset. However, because $\mathcal{G}_{r/Brexit}$ had significantly more comment-reply pairs than the other forums, we retained only the users (nodes) who commented on at least ten posts over the course of a given month. The number of nodes and edges in each graph is provided in Table 7.7. The final dataset is comprised of a total of 49,140 comment-reply pairs, forming a temporal user interaction graph for each subreddit.

Annotation

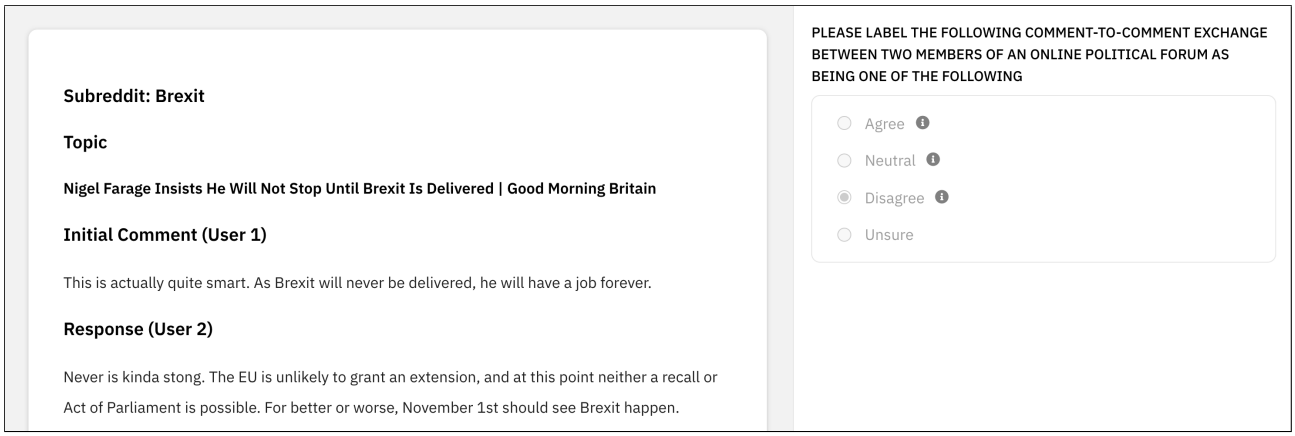


Figure 7.2: User interface for annotators

Crowdsourcing annotation setup Comment-reply pairs were annotated with *agree*, *neutral*, *disagree*, or *unsure* labels by a team of 529 English-proficient annotators from Scale AI. The annotators were provided with introductory courses on the BLM and climate change movements, as well as UK and US politics. They were also provided with examples for each of the four labels, with the reasoning behind the labels attributed.

Annotators were trained using a combination of i) an instruction document, ii) training webinars taught by Scale AI's internal subjective matter expert, and iii) training courses and quizzes based on the instructions and a sample golden dataset. The final set of annotators was selected based on their performance on a gold standard dataset of 74 comment-reply

pairs.

Each annotator was given a confidence score based on their accuracy on the gold standard dataset. Annotators were presented with a ten page 'Instructions' document, accompanying each task, as a reminder of annotation best-practices and with ten examples. Details from the instructions document are presented in the supplementary material. The annotator user interface is depicted in Figure 7.2: each task includes the subreddit name, the initial post (topic), the comment-reply pair, and a list of subreddit-specific abbreviations. Annotators are being compensated above minimum wage in their respective labor markets in order to ensure ethical and fair compensation.

Labelling and inter-annotator agreement Each task is annotated by three to five annotators based on Scale AI's dynamic consensus process. Specifically, the number of annotators per task is determined by averaging the confidence score of the annotators and comparing it against a minimal desired confidence threshold score. For each task, the majority class was decided as the final label. Tasks where the annotators equally chose *agree*, *neutral* and *disagree* were reviewed by Scale AI's internal subject experts before finalizing the response. Overall, 33% of the annotations have full inter-annotator agreement. We drop the 6,246 comment-reply pairs whose final label is unsure, and the pairs with lower than 2/3 inter-annotator agreement score, leaving a final dataset of 42,894 labeled interactions. The statistics of DEBAGREEMENT are detailed in Table 7.7.

Table 7.1: Dataset statistics

	r/Brexit	r/climate	r/BLM	r/Republican	r/democrats
Start date	Jun 2016	Jan 2015	Jan 2020	Jan 2020	Jan 2020
#nodes	722	4,580	2,516	8,832	6,925
#edges	15,745	5,773	1,929	9,823	9,624
positive	29%	32%	45%	34%	42%
neutral	29%	28%	22%	25%	22%
negative	42%	40%	33%	41%	36%

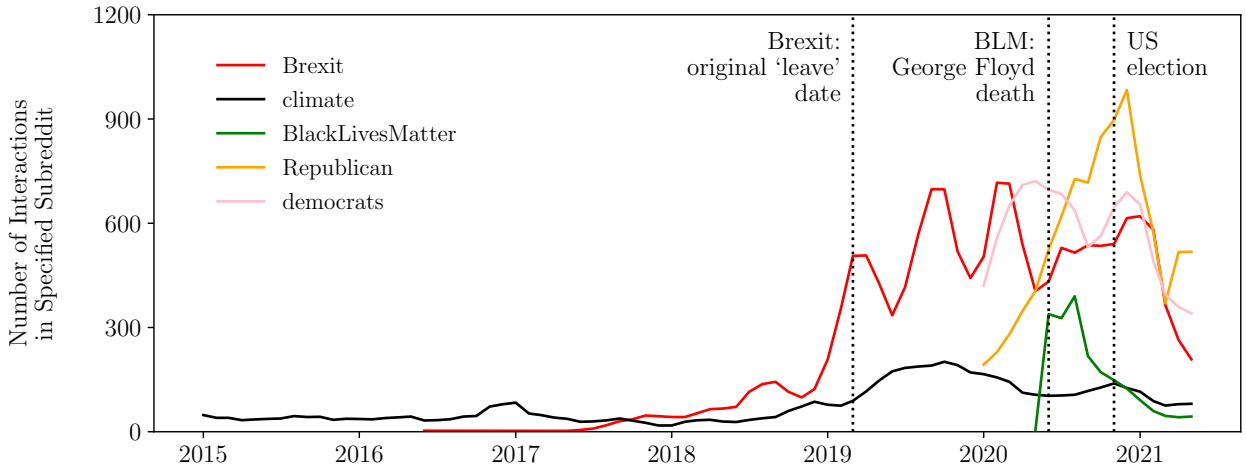
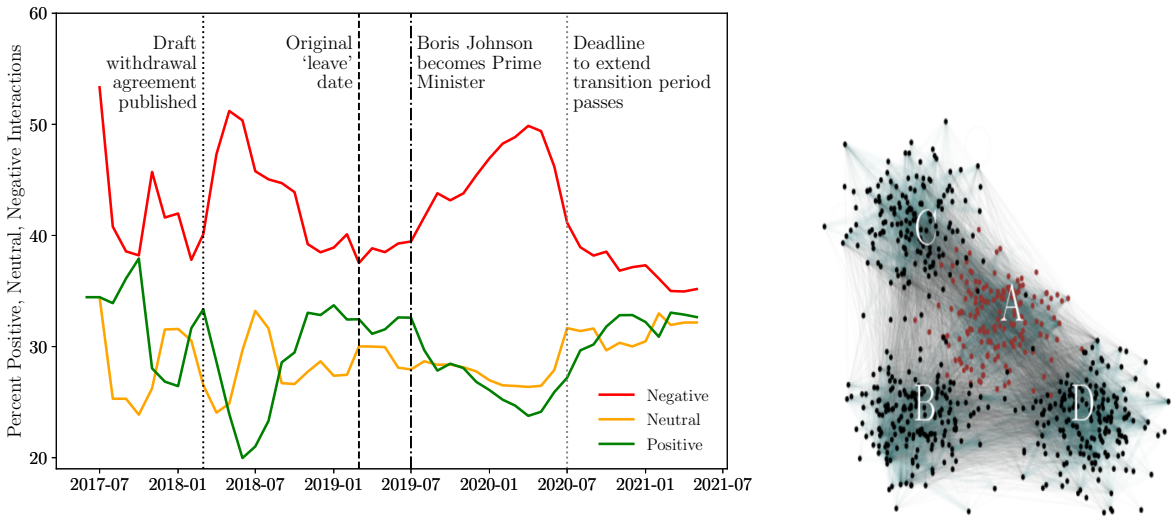


Figure 7.3: Number of interactions per subreddit (3-month rolling averages)

Activity Figure 7.3 displays how the number of interactions in DEBAGREEMENT varies over time. The activity matches key historical events in several subreddits. For example, the activity in `r/Brexit` grew dramatically as the original date of the UK's exit from the EU approached. On the other hand, `r/BlackLivesMatter` rose to prominence following the death of George Floyd. We also observe activity spikes in `r/democrats` and `r/Republican` following the 2020 United States presidential election.

Disagreements and polarisation Increased research efforts have been made to understand polarisation in social media (Tucker et al. 2018, Bail et al. 2018). In addition to offering a valuable dataset for the NLP community, DEBAGREEMENT also presents opportunities to study online social movements. This paragraph gives an overview of the insights that can be gained by using the DEBAGREEMENT dataset by taking a closer look at the annotated (dis)agreement data from `r/Brexit`.

Figure 7.4a displays how positive, negative and neutral interactions have evolved in `r/Brexit` over time. We observe that the publication of the initial draft withdrawal agreement



(a) Interactions in r/Brexit (3-month rolling averages) (b) Communities in r/Brexit

Figure 7.4: Polarisation in r/Brexit over time

brought division within the forum, with a drastic drop in the fraction of agreements and a rise in disagreements. Similarly, the election of Boris Johnson as Prime Minister and the subsequent negotiations were both correlated with a gradual rise in disagreements.

When aggregating the data over the whole time period in r/Brexit, the resulting (static) graph of discussions offers additional insight into polarisation. We mine communities of individuals with similar opinions by applying a conventional community detection algorithm for signed graphs (Traag & Bruggeman 2009) to the annotated r/Brexit data. Such algorithms aim to find graph communities by maximising the number of positive edges within communities and the number of negative edges between communities. As depicted in Figure 7.4b, we obtain four communities. By reading the comments posted by the most active users of each community, we conclude that community **A** (in brown) is pro-Brexit and communities **B**, **C** and **D** (in black) express sentiments in favor of the UK remaining in the EU. We look further at the main topics of discussion in each community and conclude that: users in community **B** are interested in the consequences of Brexit on international trade, users in community **C** discuss the accountability of UK political figures in what they consider 'a disaster' for the UK, and users in community **D** are mostly interested in UK-EU negotiations

and the votes in UK parliament.

DEBAGREEMENT is available to download at: https://scale.com/open-datasets/oxford.

7.1.3 Benchmark evaluations

Experimental Setup

All experiments were performed on two NVIDIA TITAN RTX 24Gb GPUs. We use the HuggingFace implementation of the language models we evaluate.

Evaluating SOTA pre-trained LMs

Table 7.2: LMs accuracy for (dis)agreement detection: standard deviation in parentheses

BERT			**XLNet**		**RoBERTa**		**DeBERTa**	
base/uncased	base/cased	large/uncased	base	large	base	large	base	large
62.4 % (0.6%)	61.8% (0.1%)	63.7% (1.2%)	63.6% (0.8%)	63.3% (0.6%)	63.2% (1.4%)	64.1% (1.3%)	63.0% (0.6%)	**64.1% (0.8%)**

We evaluate the performance of four state-of-the-art pretrained language models on (dis)agreement detection: BERT (Devlin et al. 2018), RoBERTa (Liu et al. 2019), DeBERTa (He et al. 2020) and XLNet (Yang et al. 2019). We build training samples by concatenating the parent sequence, the $[SEP]$ token, and the child sequence. We split the data into 80%/10%/10% train/val/test sets while maintaining the temporal order, where testing is done on the latest data. This follows a realistic setting in which one uses a trained model to perform prediction on new, incoming data (Lazaridou et al. 2021). The chosen train/val/test split prevents data leakage of terminology only used in future periods into training data. We train each model four times with different seeds. All models perform with average accuracy ranging from 62% to 64% (+22/24% above the majority class).

Failure modes Due to relatively similar model performances, we choose to focus on BERT base/uncased (b/u) to gain further insight into LM failure modes.

Table 7.3: BERT(b/u) performance statistics

	precision	recall	f1-score	support
disagree	64.3%	73.8%	68.7%	557
neutral	63.7%	44.1%	52.1%	517
agree	60.1%	69.4%	64.4%	500

Table 7.4: BERT(b/u) confusion matrix

		Predicted Label		
		-	**0**	+
True Label	-	74%	11%	15%
	0	28%	44%	28%
	+	17%	14%	69%

The model performs worst at identifying *neutral* interactions. Qualitatively, examples of *agree* and *disagree* classes show clear support or animosity between two users. On the other hand, *neutral* examples, which lack engagement and/or express partial agreement and disagreement, have the potential to confuse the model.

We present a comment-reply pair which BERT(b/u) labeled as *agreement* with high confidence, however, annotators correctly identified as a *neutral* interaction:

Subreddit: Brexit

Parent: The government don't have anything against immigrants personally. In fact they know that the economy thrives on healthy immigration. But they capitalized on a xenophobic voter base, so they have to sneak stuff like this in under the radar.

Child: Although I think you're right the government knows the UK needs immigrants I'm not so sure about their personal opinions. Mrs. May did a lot to make life harder for immigrants with her hostile environment policies. Much more than was necessary to appease the xenophobic voter base.

We observed that interactions with the largest loss values are often labeled *agree* or *disagree* by BERT(b/u), even though they are considered *neutral* by annotators. This points to

the fact that current SOTA LMs still fail to capture the subtleties of human dialogue when complex or nuanced interactions occur (Ettinger 2020, Qiu et al. 2020).

Formal vs. informal online interactions We argue that: i) current LMs struggle with messy data when it is either scraped directly from social media or pulled from human dialogue, and ii) clean and formal datasets may not have transferable insights for online (dis)agreement detection.

As outlined in section 7.1.1, most existing stance detection datasets are either small, focused on one particular topic, and/or contain formal, structured discussions. We consider the Perspectrum dataset (Chen et al. 2019), a formal text dataset for (dis)agreement detection, and compare BERT(b/u) performance on DEBAGREEMENT and Perspectrum. For consistency, we retain only *support/oppose* annotations in Perspectrum, and *agree/disagree* interactions in DEBAGREEMENT.

Table 7.5: Accuracy of BERT(b/u) - DEBAGREEMENT vs Perspectrum

	Brexit	Republicans	Democrats	Climate	BLM	Perspectrum
All subreddits	82.1%	78.4%	81.0%	83.9%	79.2%	57.7%
Perspectrum	56.4%	54.1%	57.1%	55.2%	58.4%	90.5%
Most frequent class	52.5%	52.7%	51.0%	58.3%	69.8%	52.9%

In Table 7.5, we compare the accuracy of BERT(b/u) when trained on all subreddits (first row) or Perspectrum (second row), and tested on each subreddit and Perspectrum (columns). We observe an accuracy of 90.5% when training on Perspectrum and evaluating performance on the same dataset. However, BERT(b/u) trained on Perspectrum performs poorly on DEBAGREEMENT. The fine-tuned model even fails to outperform the naive most-frequent class estimate for subreddits where the class balance is poor (`r/BlackLivesMatter` and `r/climate`). These findings imply that disagreement detection in the online, messy setting is a fundamentally different problem to the formalised, structured setting, with relatively few transferable insights between them.

Alternative training data

In this section, we consider the performance of BERT(b/u) with different choices of training data from DEBAGREEMENT. We show that i) BERT(b/u) performs better on a specific subreddit when trained on data from other subreddits, and ii) masking either of the two comments during training leads to lower performance.

Cross-subreddit evaluation We consider BERT(b/u) performance on a specific subreddit when trained on the subreddit, other subreddits, or all subreddits. In Table 7.6, we observe that training on data across different subreddits improves performance compared to training on the test subreddit alone. This suggests the capacity of LMs to learn signal beyond subreddit-specific terminology and jargon.

Table 7.6: Cross subreddit BERT(b/u) accuracy

		Subreddit				
		Brexit	Republicans	Democrats	Climate	BLM
Training data	Most frequent class	35.4%	40.3%	39.7%	41.2%	53.6%
	Current subreddit	62.4%	64.6%	65.3%	62.2%	58.3%
	Other subreddits	62.2%	64.3%	**66.9%**	65.4%	**70.8%**
	All subreddits	**64.1%**	**64.9%**	66.9%	**66.1 %**	68.6%

Masking parent and child comments We also assess the relative importance of the parent and child comments for the (dis)agreement detection task. We compared the accuracy of BERT(b/u) trained with both the parent and child comments, and trained with each one of them only, on `r/Brexit`. The baseline accuracy (when the model is trained with both comments) is 62.4%. We observed an accuracy of 38.1% on the test set when we trained the model with the parent message only, and an accuracy of 60.7% when training with the child message only. This suggests that: i) the child comment contains most of the signal required for the task, and, most importantly, ii) the parent comment provides textual context which

improves (dis)agreement prediction.

7.1.4 Limitations and Future work

Combining LMs and GRL methods Recent research has proven the benefits of leveraging contextual information in language modeling, whether it is structured knowledge (Colon-Hernandez et al. 2021) or social context (Keskar et al. 2019, Kulkarni et al. 2021). GRL methods have become increasingly popular for natural language processing (Wu et al. 2021). Colon-Hernandez et al. (2021) reviews how structured knowledge, such as knowledge graph (KG) embeddings (Wang, Mao, Wang & Guo 2017), have been combined with language models. An example is KnowBERT (Peters et al. 2019), a language model which injects KG embeddings into BERT's internal layers so the model learns knowledge-aware token representations. In another context, (Kulkarni et al. 2021) build a geographically-sensitive language model for token prediction in tweets of the form: *The most popular NFL team in my state is [MASK]*. The authors use Node2Vec (Grover & Leskovec 2016) on a graph in which nodes are US cities and edge weights are geodesic distances between cities. They show that injecting the embedding of the location of the tweet author into the internal layers of BERT improves BERT's performance.

DEBAGREEMENT presents an opportunity to train socially aware language models. For instance, one may use user embeddings trained on the DEBAGREEMENT graphs, and inject them into state-of-the art LMs. One may also consider (dis)agreement detection as a sign link prediction task on the DEBAGREEMENT graphs. In this setting, the textual information could be treated as edge or node features, and suitable Graph Neural Networks (GNN) may be applied for edge sign prediction. Social balance theories (such as the theory that *A friend of my enemy is my enemy*) may be relevant for inferring missing edges (Cartwright & Harary 1956).

Labeling challenges Despite the preprocessing we performed to ensure high-quality annotations, only 33% of the pairs were annotated with full inter-annotator agreement. This speaks to the difficulty in annotating short online exchanges containing nuanced statements, sarcasm, and diverse writing genres. Furthermore, the boundary between a neutral and a negative reply, or a positive one, may be blurred. This is underscored by the difficulty LMs have to identify *neutral* interactions.

We suggest two potential ways to combat labeling challenges through future work. One involves drawing more data from the forum for better context - specifically, we suggest potentially annotating full discussion threads, instead of comment-to-comment interactions in isolation (see below). Additionally, we see some potential in training NLP tools on interactions where annotators did not agree to identify examples that need additional attention.[3]

Comment-reply pairs vs discussion threads As discussed in section 7.1.2, `reddit` users may engage in lengthy back-and-forth exchanges, rather than a single comment-reply interaction. In DEBAGREEMENT, we consider initial comment-reply pairs as a proxy for discussion threads. By doing so, we may fail to capture some of the complexity of an entire online conversation, and opportunities to foster dialogue modeling research.

Exploring more subreddits DEBAGREEMENT follows five key socio-political movements. Expanding annotations to other areas of discussion (such as consumer products reviews, `r/buyitforlife` or wider political forums, `r/politics`) could lead to further discoveries. Annotated data from these subreddits may allow us to glean new insights about future consumer preferences or rising areas of political contention.

[3]Importantly, these findings date to research performed before the publication of OpenAI's ChatGPT (OpenAI 2023)

7.2 Temporal Signed Louvain: the clustering algorithm for a turbulent world

Political events, such as wars, changes in government and civil movements, have changed and shaped the course of history. On February 22, 2022 Russian president Vladimir Putin announced the beginning of a 'special military operation' in Ukraine – the culmination of several years of worsening relations between the two countries. In October, 2023 the world witnessed the bloody unraveling of the Israel-Palestine conflict. Simultaneously movements such as the BlackLivesMatter protests and climate change discussions continue to shape policy. Finally, the year 2024 is set to be the 'biggest global election year in history.[4] Evidence points to the fact we are living through events that will shape the trajectory of the coming decades and centuries.

As tensions continue to rise and heated debates unfold, how can researchers understand and quantify the dynamic world we live in? Network tools provide an important lens, allowing us to view the world as a complex system and study the connections between different agents. Clustering exercises in particular allow us to identify which groups of agents are most closely connected to each other and may share similar views / characteristics. In the case of social movements, clustering may allow us to identify individuals who support a similar ideology. At the international scale, we can identify which countries are potentially allied with each other.

The network science community has had a long-standing interest in efficient methods for uncovering clusters in static, unsigned graphs. Despite these algorithms' effectiveness at uncovering important characteristics of many real-world examples, studying the evolving dynamics between agents that can reach different levels of agreement / disagreement

[4] https://www.vox.com/future-perfect/2024/1/3/24022864/elections-democracy-2024-united-states-india-pakistan-indonesia-european-parliament-far-right-voting

requires a more nuanced approach. We propose a temporal, signed clustering algorithm to help capture the dynamics. Exploring such methodologies has gained momentum within the network science community in recent years – with both signed clustering algorithms and temporal clustering algorithms receiving significant attention in the literature. We discuss the problem framing for signed clustering and temporal clustering (as well as citing seminal works for important methods) in greater detail in Appendix 7.A.1; Bara'a et al. (2021), Saxena et al. (2017) are two relatively recent reviews of traditional clustering methods, while Al-Andoli et al. (2021) also consider recent deep learning techniques. Relatively few methods explore temporal, signed clustering methods – this thesis complements the evolutionary and differential temporal, signed network algorithm literature (Chen, Liu, Hao & Wang 2020, Chen et al. 2016, Wang, Song, Lu & Wang 2017).

The goal of this section is twofold: i) we present a novel flexible temporal, signed clustering algorithm and ii) we present insights from applying it in several topical settings. Specifically, we study generalisations of the Louvain method (Blondel et al. 2008), a popular and efficient method for modularity optimisation in unsigned graphs. We propose a new heuristics-based method, Signed Louvain, to maximise the signed modularity function (Traag, Doreian & Mrvar 2019). We then expand the heuristics to temporal signed graphs. The thesis applies the novel method in two settings: i) to study the five online discussion communities annotated within DEBAGREEMENT discussed earlier in this chapter in order to uncover the dynamics behind political movements (similar in spirit to Grindrod & Bovet (2022)), as well as to study the evolution of international relations from the perspective of international news events from Integrated Crisis Early Warning System (ICEWS) dataset (Boschee et al. 2015).

7.2.1 Methods

Notation We consider a temporal network represented as a stream of time-indexed, undirected, signed edges $\{e_t\}$ appearing over continuous time t. An edge is denoted by the following tuple $e_t = (t, u, v, w^-_{t,u,v}, w^+_{t,u,v})$, where t is the time at which the edge appears, u and v are unique node identifiers (importantly, nodes can be added to the network over time and receive a unique identifier once added to the network). $w^-_{t,u,v}, w^+_{t,u,v}$ represent the positive and negative weight associated with the edge. If, for example, e_t represents a fully positive interaction between nodes u, v then $w^-_{t,u,v} = 0$ and $w^+_{t,u,v} = 1$. The framework allows us to incorporate a continuous spectrum of interactions; we choose to quantify the positive, $w^+_{t,u,v}$ and negative, $w^-_{t,u,v}$, interactions separately for computational purposes.

To evaluate the communities of the network at time t, we aggregate all edges up to time t into the adjacency matrix $A(t)$, which is the sum of the negative and positive adjacency matrices of the network: $A(t) = A^+(t) - A^-(t)$.

The importance of an edge in the evolving network decreases with time. Therefore, the weight of edges that occur before time t is decayed exponentially. For two given nodes u and v, their entry in the temporal adjacency matrix is computed as:

$$A^+_{uv}(t) = \sum_{t_k \leq t} w^+_{t_k,u,v} \exp(-\eta(t - t_k)), \tag{7.1}$$

$$A^-_{uv}(t) = \sum_{t_k \leq t} w^-_{t_k,u,v} \exp(-\eta(t - t_k)). \tag{7.2}$$

Given this network specification, we design Temporal Signed Louvain (TSL), our method to detect communities in temporal and signed networks.

Algorithm Design

Modularity Maximization – Unsigned, Static Method As a starting point, we use the Louvain method (Blondel et al. 2008), a modularity maximisation-based method for community detection in unsigned, static networks. The modularity of a network partition is a scalar value between −1 and 1 that measures the density of links (edges) inside communities as compared to links between communities. Based on the homophily assumption, the intuition is that a well partitioned network should have most links inside communities and not between communities. Given a weighted network with adjacency matrix A, the modularity of partition σ is defined as:

$$Q(\sigma) = \frac{1}{2m} \sum_{u,v} \left(A_{uv} - \gamma \frac{k_u k_v}{2m} \right) \delta(\sigma_u, \sigma_v) \tag{7.3}$$

where A_{uv} is the edge weight between u and v, $k_u = \sum_x A_{ux}$ is the degree of node u, σ_u is the community to which node u is assigned, γ is a resolution parameter, and δ-function is the indicator function. Finding the partition that maximises Equation (7.3) is a NP-hard problem; so the Louvain method (discussed below) uses an efficient heuristics that approximate the best solution.

Modularity Maximization – Signed, Static Method Following its success, modularity-maximisation methods have been explored for signed networks. In Traag & Bruggeman (2009), a signed modularity function is defined as follows. The network is split into its positive and negative parts such that the signed adjacency A can be written $A = A_+ - A_-$ and the modularity of a signed network can be defined $Q = Q_+ - Q_-$ following the definition in

Equation (7.3) on the positive and negative network parts respectively:

$$Q(\{\sigma\}) = \frac{1}{2m}\sum_{u,v}\left(A_{uv} - \left(\gamma^{+}\frac{k_u^{+}k_v^{+}}{2m^{+}} - \gamma^{-}\frac{k_u^{-}k_v^{-}}{2m^{-}}\right)\right)\delta(\sigma_u, \sigma_v) \tag{7.4}$$

where A_{uv} is the edge weight between u and v, $k_u^{+} = \sum_x A_{ux}^{+}$ is the positive degree of node u (total weight of positive edges linking to node u), $k_u^{-} = \sum_x A_{ux}^{-}$ is the negative degree of node u (total weight of negative edges linking to node u), σ_u is the community to which node u is assigned, γ^{+} is the positive resolution parameter, γ^{-} is the negative resolution parameter, and δ-function is the indicator function. This definition of modularity effectively rewards placing nodes with many positive links into the same community, but penalizes placing nodes with many negative links into the same community.

Modularity Maximization – Signed, Temporal Method (TSL) To create the TSL, we starting with an initial non-empty network (snapshot) at $t = 0$, we detect initial communities leveraging the Reichardt & Bornholdt (2006) approach to compute the positive and negative modularities separately, and a `Louvain Optimizer` multiplex partitioning to compute the overall optimal partitioning. Our approach offers flexibility to utilize other algorithms – depending on the application, a different optimization methodology may be more appropriate and our approach is flexible to incorporate such modifications. For our approach, we test the Louvain method, whcih is a hierarchical clustering algorithm that recursively merges communities into a single node and re-evaluates / executes clustering on the condensed graphs, summarized in Algorithm 1.

The configuration allows for several important choices. In line 4 above, the algorithm traditionally looks to merge nodes with neighboring communities. In our case of signed graphs (where links can indicate heterophily as well as homophily), we elect to consider all communities for the node move. The second option is around the choice of quality

Algorithm 1: Louvain Outline

```
Initialise all nodes in their own independent communities;
while Quality function improvement > 0 do
    for Node in Graph G: do
        Compute maximum quality function improvement for moving node to a
         community within the graph;
        if Maximum quality function improvement > 0 then
            Merge node into the community which maximizes the quality function. In
             the updated graph, the node and community will be represented by a
             single node (which inherits the links of all subsumed nodes);
        else
            Continue
        end
    end
    return Updated graph G
end
```

function. We leverage the `louvain.RBConfigurationVertexPartition`[5], which computes a non-normalized version of modularity (treating the positive and negative graphs are separate layers of the same graph):

$$Q^+ = \sum_{u,v} \left(A^+_{uv} - \gamma^+ \frac{k^+_u k^+_v}{2m^+} \right) \delta(\sigma_u, \sigma_v),$$
$$Q^- = \sum_{u,v} \left(A^-_{uv} - \gamma^- \frac{k^-_u k^-_v}{2m^-} \right) \delta(\sigma_u, \sigma_v).$$

The algorithm then maximizes modularity for the positive and negative layers simultaneously by computing the modularity as: $Q = Q^+ - Q^-$.[6]

Once an initial community allocation has been set, the timestamps edges are processed one at a time. Each time a new edge appears (or at predefined intervals) communities are recomputed: the algorithm is initialized with the previous time step's community allocations; the method for modularity optimization from Reichardt & Bornholdt (2006) in combination with `Louvain Optimizer` multiplex partitioning is then applied to the updated graph G_t to update node allocations.

[5] https://louvain-igraph.readthedocs.io/en/latest/reference.html
[6] https://louvain-igraph.readthedocs.io/en/latest/reference.html

Previous works in community detection have focused on detecting communities either in static signed networks or in temporal unsigned networks. TSL fills an important gap in the literature by formulating the first modularity-maximizing community detection algorithm for temporal signed networks. It has several advantages relevant for social dynamics analysis that we detail below. Because of its ability to detect community changes continuously, we can naturally identify significant structural changes in the structure of online networks. Given the flexible aggregation function, the algorithm can also easily be extended to include other classes of interactions (such as multilayer interactions in contexts where users interact across several dimensions).

Hyperparameters There are several hyperparameters that need to be set in TSL. γ^+ and γ^- in Equation (7.3) are typical resolution parameters of Louvain-inspired methods. In our application, careful consideration of γ^+ and γ^- is of particular importance as they play an important role in determining community size and connectedness in sparse graphs. We discuss their importance in greater details, as well as the parameter choices in this setting in Appendix 7.A.2.

η in Equations 7.1, 7.2 is the tie-decay parameter. It controls for how fast the weight of past edges collapses to zero. A higher tie decay gives more importance to recent interactions and makes older ones obsolete. This parameter is set via parameter search by choosing the values maximising stability of communities across several runs of the method. Eventually, because past edges' weights decrease to zero, nodes which have not been active for a certain period of time become isolated and create small, noisy network components disconnected from the rest of the network and uninformative for the community inference process. For this reason, we introduce an activity threshold ζ so that any node whose total activity weight is smaller is simply removed from the network before community inference. For interpretability, we set η as the fraction of an original interaction's weight remaining

after three months time. If η is set to 0.3 – 30% of the interaction's weight will remain in three month's time. If η is set to 0.9 – 90% of the interaction's weight will remain in three month's time. If η is set to 1 – the algorithm will work with no tie decay; new information will be weighted equally to old information. We explain in greater detail the methodology and significance of η in Appendix 7.A.2.

Additionally, we observe that the stability of the method is dependent on network density. For this reason, we proposed several methods to clean and select highly-active, community-defining nodes prior to running the algorithm. Our code base allows one to filter data based on a global clustering coefficient, ensuring that data remains connected over time and to propose potentially unobserved links due to triadic closure.

Synthetic Experiments

Evaluation framework Our synthetic experiments use the following general evaluation framework. We begin by generating signed network data where node communities are known a-priori, and data is generated to match certain edge densities within and across communities – a method referred to as a Signed Stochastic Block Model (SSBM) (Holland et al. 1983, Jiang 2015). We use the Normalized Mutual Information (NMI) as our metric of choice to evaluate performance of the TSL (Strehl & Ghosh 2002). This metric evaluates how close the communities found by TSL are to the ground-truth communities defined as part of the data generation process.

NMI ranges between zero and one, with a score of one signifying a perfect match of assignments, and has become standard for evaluating the quality of community detection algorithms (Fortunato & Hric 2016). Specifically, if we have N objects, and two label assignments U,V of these objects. The entropy a specific partition, where we use i to index

communities in the summation below, is:

$$H(U) = -\sum_{i}^{|U|} P(i)log(P(i)),$$

where $P(i) = |U_i|/N$ is the probability of selecting a community of size $|U_i|$ at random from the N items. The entropy measures the amount of uncertainty associated with each partition. Therefore, if a particular partition is likely to occur by randomly placing objects into communities, its entropy is high.

The mutual information between two partitions, U, V is:

$$MI(U,V) = \sum_{i}^{|U|}\sum_{j}^{|V|} P(i,j)log\left(\frac{P(i,j)}{P(i)P(j)}\right),$$

where $P(i,j) = |U_i \cap V_j|/N$ (the probability for objects in communities U_i and V_j to intersect randomly).

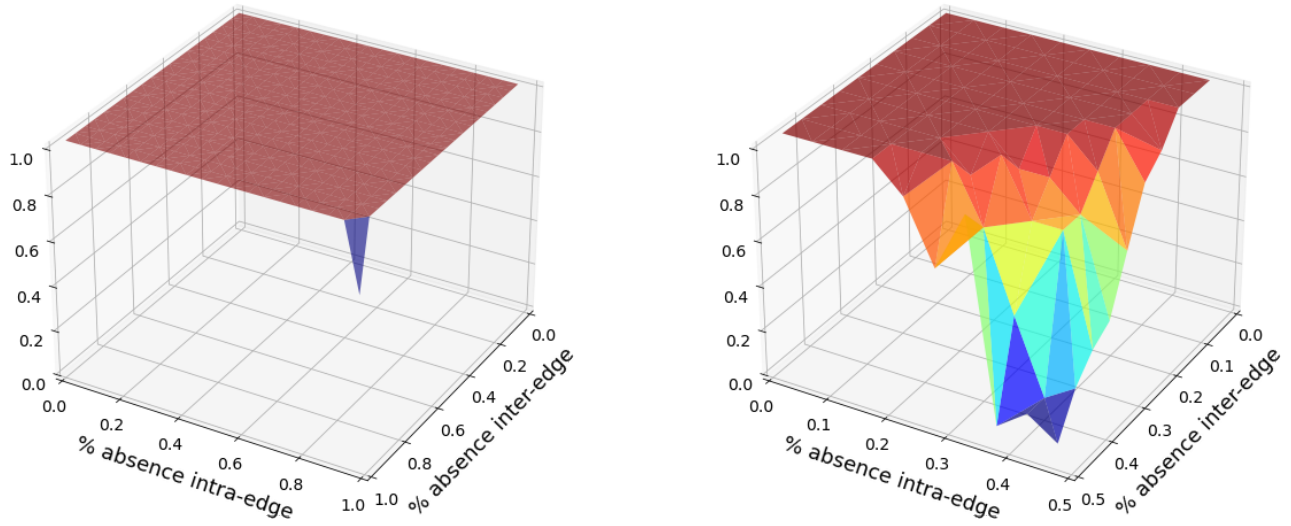
From these two components, the NMI between two partitions U, V is defined as:

$$NMI(U,V) = \frac{2MI(U,V)}{H(U)H(V)} \tag{7.5}$$

Static Experiments

We consider the performance of the algorithm in two static settings: one where we vary the fraction of edges in the network, and another where we vary the noise in terms of fraction of positive versus negative links between and within communities, respectively. The first scenario is intended to test robustness to network sparsity. The second looks at the performance of the algorithm in the presence of signal noise within and between communities, where not all users within a community agree with each other and disagree with other members.

The results with respect to sparsity are presented in Figure 7.5a. This scenario has no



(a) Algorithm Robustness to Network Sparsity

(b) Algorithm Robustness to Noise

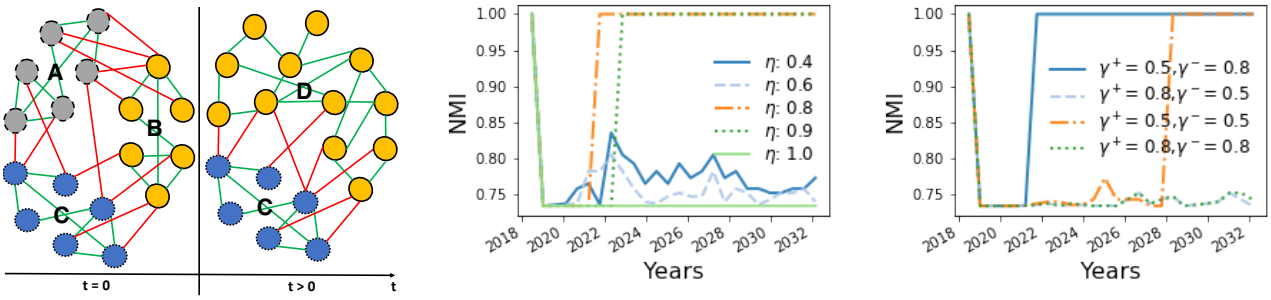
Figure 7.5: **Algorithm Robustness for Static Data:** we detect communities in synthetically-generated static network data. The chart on the left captures our algorithm's robustness to sparsity as we vary the fraction of edges missing within and between communities; all links within communities are positive and between communities are negative. The chart on the right captures the robustness to noise as we vary the fraction of positive edges between communities and negative edges inside communities, while fixing the fraction of observed edges between nodes at 0.7 (30% of edges missing). The algorithm is evaluated by the similarity of detected communities with those used in the data-generation process using the NMI metric for community comparison. In the experimental setup, we choose $\gamma^+ = 0.5$, $\gamma^- = 0.8$ and re-run the optimization while varying the random seed 100 times per scenario and keeping the community with the highest detected gain in the objective function.

noise, i.e. all links between nodes within communities are positive, while links between nodes in different communities are negative. We observe that the algorithm performs well in most regions of the parameter space. Edges missing between communities appears to present a challenge in regions where the fraction of intra and inter community edges missing is over 95%. This performance is related to our choice of hyperparameters γ^+ and γ^-, as discussed in Appendix 7.A.2.

The results with respect to noise are presented in Figure 7.5b. We vary the fraction of edges within communities that are negative between zero and 0.3. We do the same for positive edges between communities. These scenarios are meant to mimic true social dynamics where, for example, the majority of people in a group may agree with each other, but some disagreements are present. We choose the 0.3 maximum cutoff for noise as any greater

amount of noise results in communities that are inherently not well-defined. We uniformly choose fifty percent of possible edges between nodes to be present, placing us in a region of good algorithmic performance with respect to sparsity, as shown in Figure 7.5a. The results suggest that our algorithm is fairly robust to noise, with performance deteriorating only in the highest noise region where over 25 % of edges within and across communities flip sign. The results in the static setting suggest that the algorithm is sufficiently robust to noise to potentially achieve reasonable performance in the temporal setting.

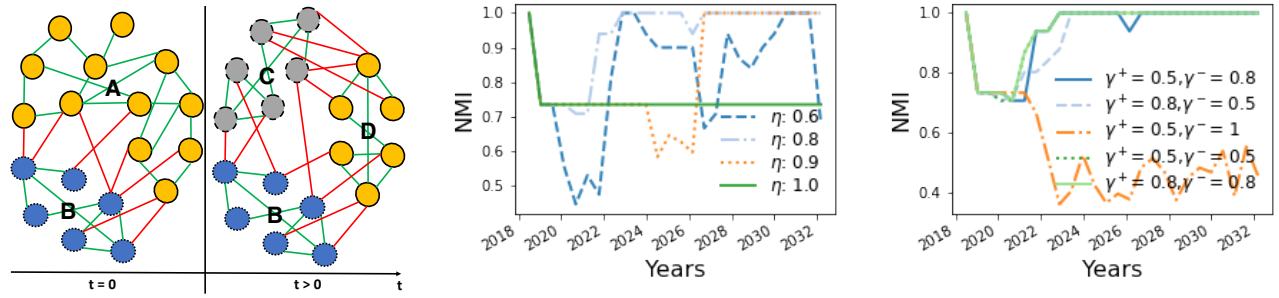
Dynamic Experiments



(a) Community Merge Diagram (b) η Experiments (c) γ experiments

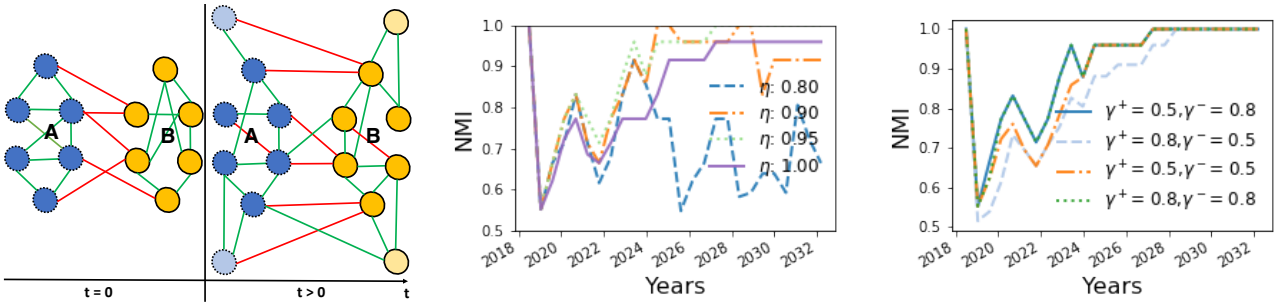
Figure 7.6: **Community Merge Experiments**. Figure 7.6a depicts a toy diagrams for the scenario. Positive (negative) edges are depicted in green (red). Figures 7.6b, 7.6c depict the ability and speed of TSL to detect community structure changes, as measured by comparing algorithm communities to the ground truth from the data generation process using the NMI metric. In Figure 7.6b we set $\gamma^+ = 0.5$ and $\gamma^+ = 0.8$ for all runs. In Figure 7.6b, we set η to 0.8.

We modify the Stochastic Block Model framework (Holland et al. 1983, Jiang 2015) to our temporal setting in order to investigate the influence of tie-decay η and the scaling parameters γ^+, γ^- on the ability of TSL to detect three temporal community dynamics: a community merger, a community split and new nodes entering existing communities. The scenarios are depicted in Figures 7.6a, 7.7a, 7.8a. In addition to testing the TSL in these settings and discussing the choice of hyperparameters, we provide a flexible framework temporal Stochastic Block Model framework, which enables one to generate synthetic data for any of the scenarios discussed while varying the amount of noise and network density.



(a) Community Merge Diagram (b) η Experiments (c) γ experiments

Figure 7.7: **Community Split Experiments**. Figure 7.7a depicts a toy diagrams for the scenario. Positive (negative) edges are depicted in green (red). Figures 7.7b, 7.7c depict the ability and speed of TSL to detect community structure changes, as measured by comparing algorithm communities to the ground truth from the data generation process using the NMI metric. In Figure 7.7b we set $\gamma^+ = 0.5$ and $\gamma^+ = 0.8$ for all runs. In Figure 7.7b, we set η to 0.8.



(a) New Nodes Diagram (b) η Experiments (c) γ experiments

Figure 7.8: **New Node Experiments**. Figure 7.8a depicts a toy diagrams for the scenario. Positive (negative) edges are depicted in green (red). Figures 7.8b, 7.8c depict the ability and speed of TSL to detect community structure changes, as measured by comparing algorithm communities to the ground truth from the data generation process using the NMI metric. These experiments introduce noise – a fraction of edges between communities may be positive and those within communities may be negative. In Figure 7.8b we set $\gamma^+ = 0.5$ and $\gamma^+ = 0.8$ for all runs. In Figure 7.8b, we set η to 0.8.

Scenario Description Figure 7.6a shows a merge between two communities: communities **A** and **B** join together to form community **D**. The networks has an initial ($t = 0$) community configuration: the network has 60 nodes split into three equally-sized communities (**A**, **B** and **C**), and 500 randomly selected edges are generated between the communities with only positive edges within communities and only negative edges between communities. At $t = 1$, **A** and **B** merge to create community **D**, and each edge created over the course of the next 500 time steps are in keeping with this new configuration: at each time step, two nodes are chosen at random and a negative edge is created between them if they are in different

communities (i.e. one node is in community **D** and the other is in **C**), and positive if they are both in the same community (**C** or **D**). In all scenarios time periods are converted to calendar timestamps for effective processing by our temporal algorithm.

The experimental design for community splitting is similar to the community merger and is displayed in Figure 7.7a: we initialise the network with 60 nodes, two communities with 40 and 20 nodes respectively, and with 500 randomly chosen edges between them. Over the course of the next 500 time steps, new edges are created randomly in the network, according to the new configuration with three equally sized communities, where the larger community has split into two smaller ones.

The last scenario simulates a network growth. The initial network has 60 nodes split into two equally-sized communities, with 500 randomly sampled edges between them; edges are positive within communities and negative across communities. At $t = 1$, we introduce 20 new nodes with ten of them assigned to each of the two communities, with one interaction per new node. An element of noise is introduced to the experiment: the edges coming from the new nodes are positively connected to nodes within their community (and negatively connected to nodes outside their community) with 90% probability. This is displayed in Figure 7.8a with the existence of a few positive (negative) edges between (within) communities **A** and **B**. Starting at $t = 1$, we observe 500 more time-steps with an edge being generated at each time-step by selecting two nodes randomly and with a 10% probability of having a noisy sign.

TSL Performance We evaluate the performance of the TSL by comparing the detected communities across time to those of the data generation process using the NMI metric for evaluation. In the initial timestep, we are effectively measuring the TSL's ability to detect the original community structure – subsequently, we consider how long (how much new data the TSL must be exposed to with different parameter settings) before the algorithm

detects and updates the community structure.

The first set of experiments tests for the impact that the time decay parameter η has on community detection (we keep the γ^+ and γ^- constant at 0.5 and 0.8, consistently with our discussion in Appendix 7.A.2). In Figure 7.6b, we observe that $\eta = 0.9$ and $\eta = 0.8$ both allow the algorithm to detect the new community structure – a lower η of 0.8, in fact allows the algorithm to detect the change sooner and by observing fewer new datapoints. However, a very low choice of η of 0.4 or 0.6 does not allow the algorithm to detect the new community structure – the information from ties decays too quickly and the algorithm no longer has a substantial base of connections to use to make inference. Instead of detecting meaningful changes in community structure, the algorithm picks up noise from one time step to the next and updates communities based on it. On the flip side, a high choice of η is also sub-optimal - setting η to 1 results in the algorithm not updating the existing community structure since the new information it receives is insufficient to overcome the information from the initially created community structure. We observe similar insights from the Community Split and New Node scenarios. In the Community Split experiments, we also observe that the algorithm is quite noisy in the detected communities when η is too low, and fails to update to the new community structure when η is too high (Figure 7.7b). In the New Node scenario with noise, a higher η ensures that the algorithm is not overly sensitive to noisy links and maintains community cohesiveness (Figure 7.8b).

The second set of experiments considers various parameter choices for γ^+ and γ^-. In the New Node scenario, the combination of γ^+ and γ^- parameters determines how quickly the algorithm converges to the correct underlying structure. In other instances, γ^+ and γ^- play a more important role. In the Community Merge scenario in Figure 7.6c, we observe that when γ^+ is set too high (when γ^+ is set to 0.8), the algorithm fails detect a merge continues to detect three communities. This is consistent with Appendix 7.A.2 where we discuss that

a higher γ^- relative to γ^+ parametrizes the algorithm to combine sparsely connected nodes into larger communities. In the Community Split experiments, we present a scenario with the opposite problem: when $\gamma^- = 1$ and $\gamma^+ = 0.5$ the algorithm incorrectly groups smaller communities into larger ones.

Overall, our experiments highlight the ability of the TSL to dynamically pick up changes in community structure that occur over time. The experiments also underscore the importance of applying domain knowledge (such as insights into whether we expect few, sparsely connected communities or many, densely connected ones) and testing for algorithm stability when applying it to real-world data.

7.2.2 Tracking International and Political Communities using the TSL

We apply the TSL in two different settings: to individual user agreement-disagreement data from Reddit (Pougué-Biyong et al. 2021), and to ICEWS international event data. The exercise demonstrates the versatility of the TSL and allows us to draw important conclusions about how these different networks have evolved and the underlying dynamics that have unfolded.

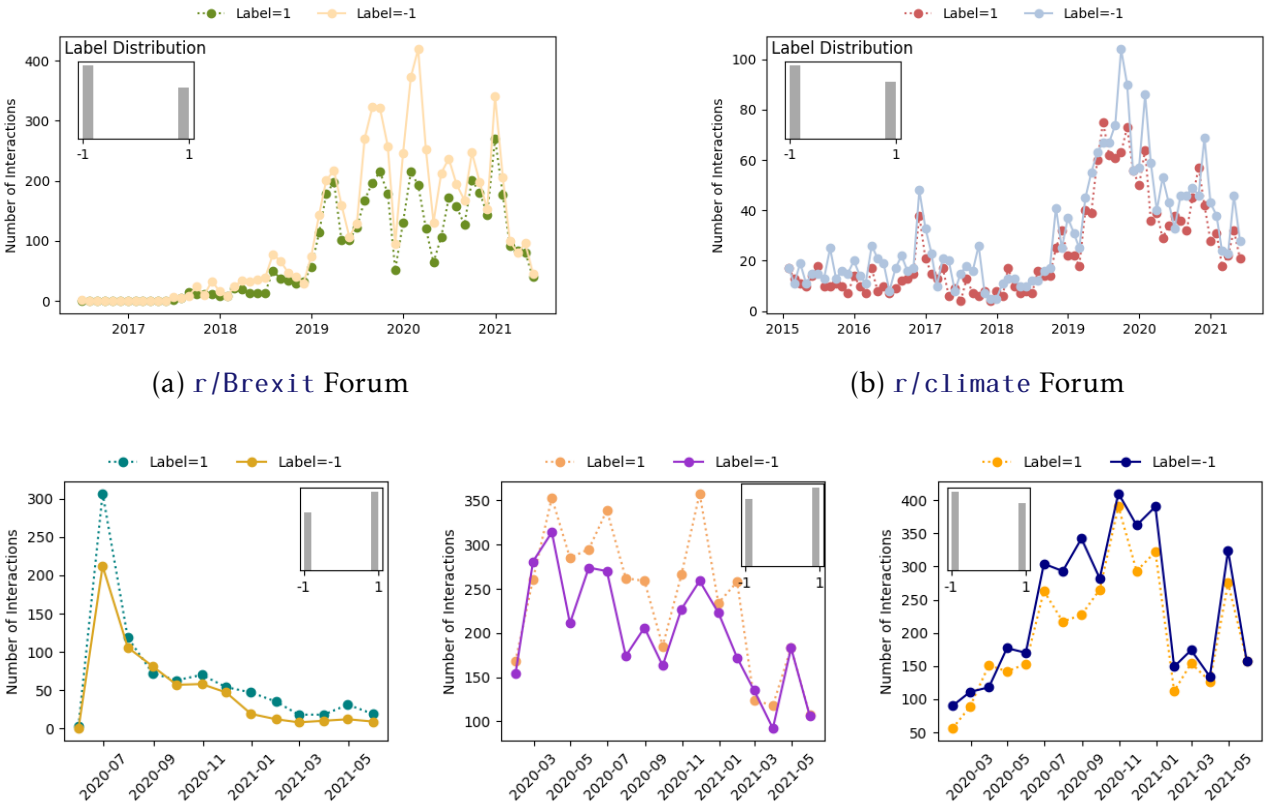
Social Media Discussions

Data Description We leverage the DEBAGREEMENT dataset (Pougué-Biyong et al. 2021) for our analysis of how political discussions among users have evolved online. This dataset was created from anonymous discussion data scraped from the Reddit platform. Reddit is a social forum broken down into smaller topic-specific discussion groups, called *subreddits*. DEBAGREEMENT contains data from five different subreddits: `r/BlackLivesMatter`, `r/Brexit`, `r/climate`, `r/democrats`, `r/Republican`. Each subreddit contains user-generated posts which are the central points of discussion: the posts can be upvoted and downvoted by other users who can also comment the posts. Users engage in topic-focused discussions

with each other in a thread format. DEBAGREEMENT data is composed of the text from posts as well as a comment-reply pairs among different users, creating a social interaction graph. The comment-reply data is annotated by trained specialists as being either *agreement*, *disagreement* or *neutral* interactions. Each interaction is annotated by at least three different annotators. The final label, as well as the inter-annotator agreement, is contained in the final dataset.

Each DEBAGREEMENT subreddit has distinct network characteristics. First of all, the annotations span different time periods. `r/BlackLivesMatter`, `r/democrats`, `r/republican` span the months preceding and following the 2020 presidential election in the United States. `r/climate` and `r/Brexit`, on the other hand, follow longer-term social movements over several years. Figure 7.9 highlights the different levels of activity across the forums across time. From a networks perspective the forums are similarly distinct in Table 7.7. For our signed, temporal clustering methodology we consider primarily negative and positive interactions: most forums have a relatively balanced number of positive and negative interactions, however, `r/climate` and `r/Brexit` contain substantially more negative than positive interactions. For all of these reasons, even though our general algorithmic framework for findings temporal communities remains the same, we must make distinct parameter choices for each of the forums and leverage our algorithms flexibility in order to select the appropriate configuration. Fortunately, the flexibility of the TSL allows us to adjust it and function well with noisy interactions from social media.

TSL Application In our experiments, we drop neutral interactions and filter out nodes with clustering coefficient equal to zero. We use our domain knowledge as well as synthetic experiments to select $\gamma^+ = 0.5$ and $\gamma^- = 0.8$. We select our time decay parameter η by comparing the NMI among the communities in which active nodes are placed into when updating community allocations using the TSL algorithm at a monthly time horizon – the



(a) r/Brexit Forum

(b) r/climate Forum

(c) r/BLM Forum

(d) r/democrats Forum

(e) r/republican Forum

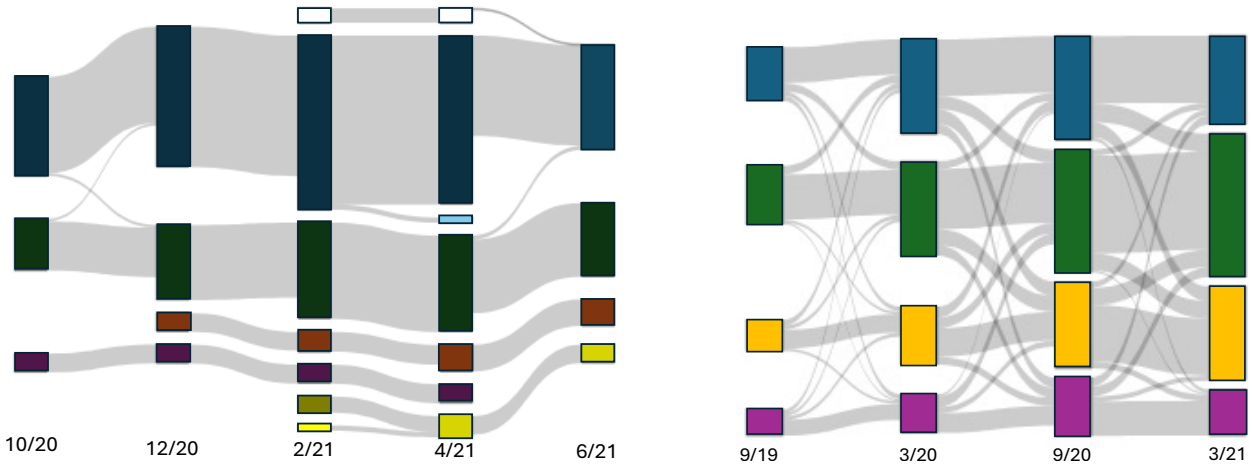
Figure 7.9: **Forum Dynamics Over Time**.

objective is to select an η parameter that allows us to detect meaningful community changes, but also one that is resilient to noise. Our experiments are presented in Appendix 7.A.3 and final η parameter choices are displayed in Table 7.7.

Beyond community detection, the TSL presents several advantageous ways to track temporal changes in community structure over time. One of the outputs of the algorithm is an alluvial diagram tracking flows from one community to another – the user has the ability to set a minimum threshold for communities to track in order to avoid looking at noisy, small communities. The alluvial diagrams allow us to observe different trends and changes across the various forums. In Figure 7.10, we observe that r/BlackLivesMatter and r/Brexit share certain common characteristics. Namely, both forums appear to have key discussion clusters that gain momentum over time. In r/Brexit, the groups appear to have greater movement with nodes changing discussion areas and political views. In

	r/Brexit	r/climate	r/BLM	r/Republican	r/democrats
Start date	Jun 2016	Jan 2015	Jan 2020	Jan 2020	Jan 2020
End date	Jun 2016	Jan 2015	Jan 2020	Jan 2020	Jan 2020
Network Characteristics					
#nodes	717	1,308	459	5,054	4,786
#edges	11,119	1,788	465	6,035	6,862
Parameters					
Data Filtering	Cluster. Coef.	Cluster. Coef.	Global	Global	Global
η	0.7	0.6	0.4	0.3	0.3

Table 7.7: **Dataset Statistics and Algorithm Parameters**; we present network characteristics across different discussion forums, as well as parameter choices. In the case of r/Brexit and r/climate we select nodes with a clustering coefficient of at least zero; in the case of r/BLM, r/Republican, r/democrats we have fewer observations and less densely connected nodes – we filter nodes based on whether they are connected to the giant component by the final time step.



(a) r/BlackLivesMatter (b) r/Brexit

Figure 7.10: **r/BlackLivesMatter and r/Brexit Alluvial Diagrams**: we observe the changes in community structure over time as individuals move between clusters and create new ones on the r/BlackLivesMatter and r/Brexit forums. For the diagrams: for r/BlackLivesMatter we set a minimum community size of ten; for r/Brexit we set a minimum community size of five.

r/BlackLivesMatter, on the other hand, even though two key discussion groups persist across time, other, smaller communities appear and die out.

The r/cliamte forum exhibits quite different dynamics. In Figure 7.11, we observe a community split in mid-2019, as well as the merger of several smaller communities over

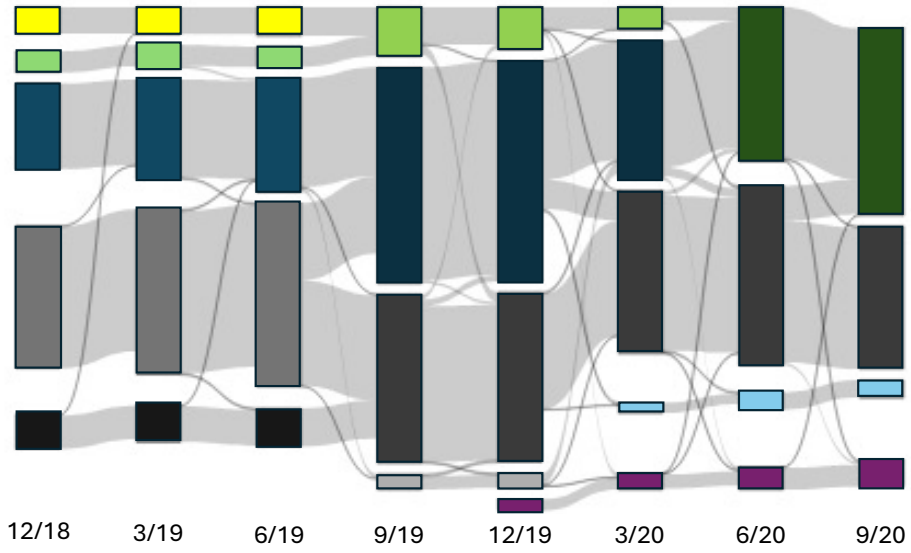
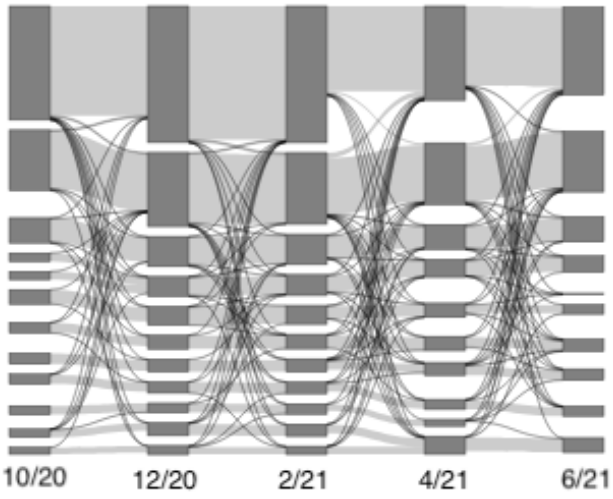


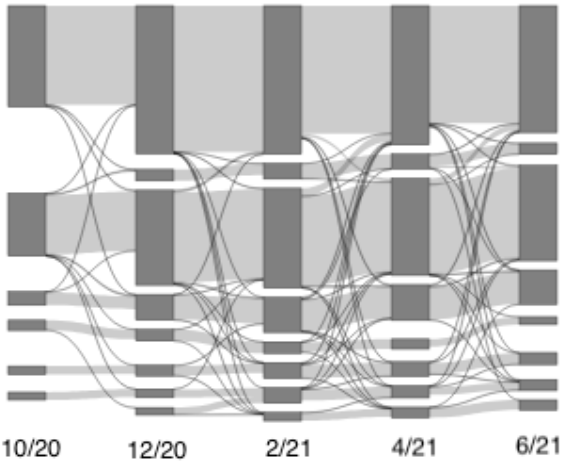
Figure 7.11: r/climate **Alluvial Diagram**: we observe the changes in community structure over time as individuals move between clusters and create new ones on the forum. For the diagrams we set a minimum community size of 15.

time into one. This trend appears consistent with the climate movement becoming more united in their political agenda and goals over time.

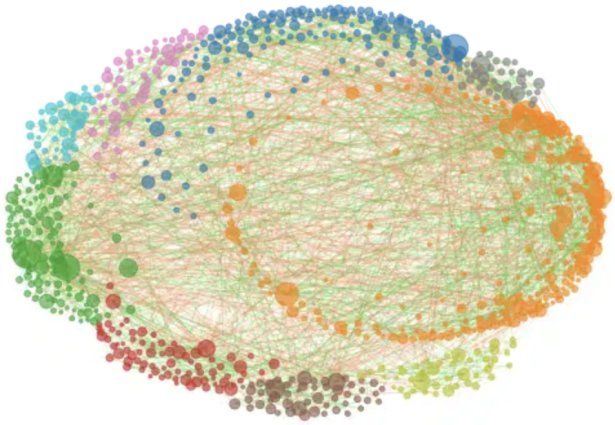
We use the alluvial diagrams in Figure 7.12 to compare the behaviour of participants in the r/democrat and r/republicans forums during the aftermath of the 2020 presidential elections. The algorithm is run with the same parametrization across both forums and alluvial diagrams are produced using the same plotting specifications. However, key differences are observable across the forum. In r/democrat, we observe substantial more movement and interchange between the different clusters. Individuals appear to move freely between different groups and engage in various conversations on the forum. In r/republicans, on the other hand, we observe fewer, larger clusters that have relatively less movement between them over time. A closer inspection of the characteristics of the two forums demonstrates that the clusters are largely formed around vocal, politically opinionated individuals that are active on the forums. However, within r/republicans individuals appear to side with a single user and keep to their opinions over time. On r/democrat active individuals also gain a following, however, forum participants often engage with a greater diversity of content produced by different users. We create a snapshot of the networks and present them in



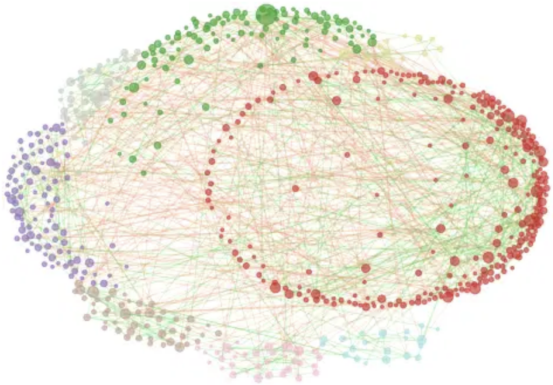
(a) r/democrat Alluvial Diagram



(b) r/republican Alluvial Diagram



(c) r/democrat Community Structure 2/21



(d) r/republican Community Structure 2/21

Figure 7.12: **r/democrat and r/republicans Alluvial and Network Diagrams**: we observe the changes in community structure over time in the r/democrat and r/republicans forums in the aftermath of the 2020 presidential election. For the alluvial diagrams for both forums we set a minimum community size of 50. For our network diagrams, we look at a snapshot of the most active users and their communities in February of 2021. Nodes are colored according to their communities and sized proportional to their degree. Red lines denote negative edges, while green denote positive ones.

Figures 7.12c and 7.12d. Both diagrams look at only users with a degree greater than four and scale nodes proportionately to their degree. On the democrat forum, we observe that user clusters are more densely connected with a mix of positive (green) and negative (red) links between communities. Within the republican forum, communities appear more polarized with many more negative links between them. The network in general is less dense with few

International Relations

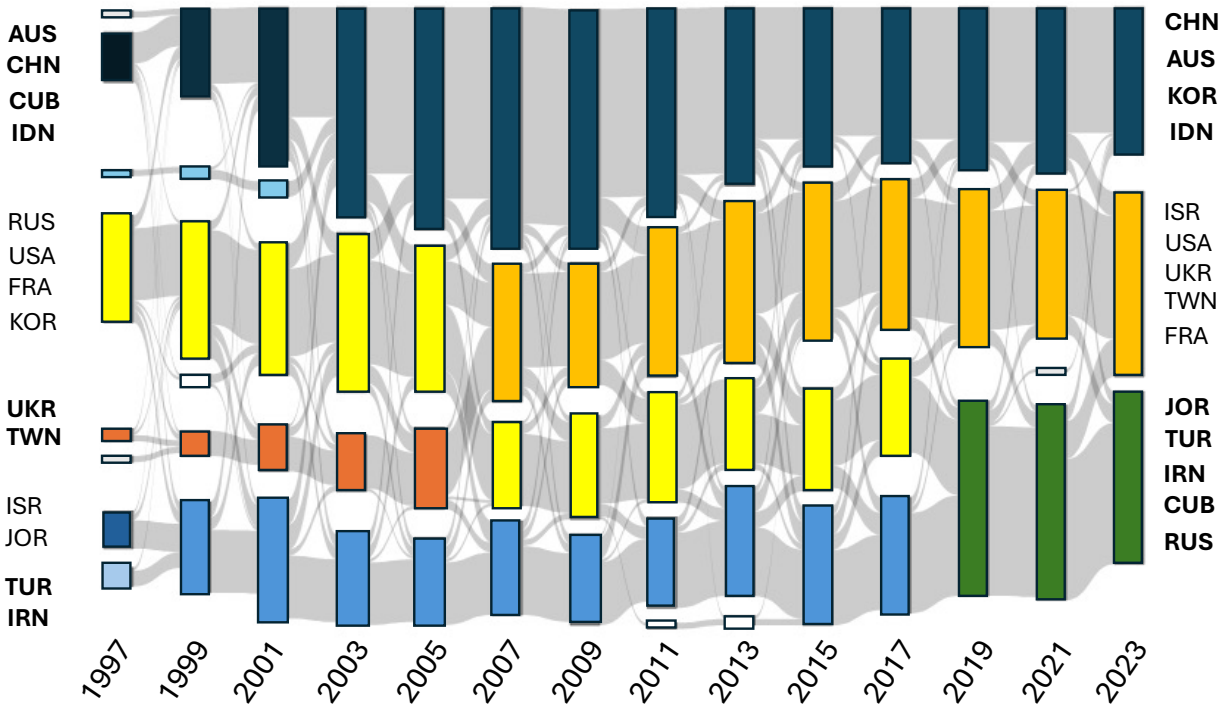
Data Description We proxy real-time developments in international relations by studying news shared within one country about another from the ICEWS dataset (Boschee et al. 2015). The key variables of interest for this study are: *Event Date*, *Source Country*, *Target Country*, *Intensity* and *Conflict and Mediation Event Observations (CAMEO) code*. We explain several of the variables below.

- *Source Country*: country in which news is created
- *Target Country*: country about which news is being disseminated
- *Intensity*: score ranked between (-10,10) documenting the direction and importance of a piece of news. A large, negative number, therefore, corresponds to a piece of news contributing to the deterioration of relations between two countries.
- *CAMEO Code*: describes the event type being discussed within the news. Events can be encoded with two to four digit codes, with four digit corresponding to greater specificity of event type. Consider for example the *Consult* (02) category, which includes specific codes for events that are identified as mediation and negotiation such as: "Engage in mediation" (025), "Engage in negotiation" (026).

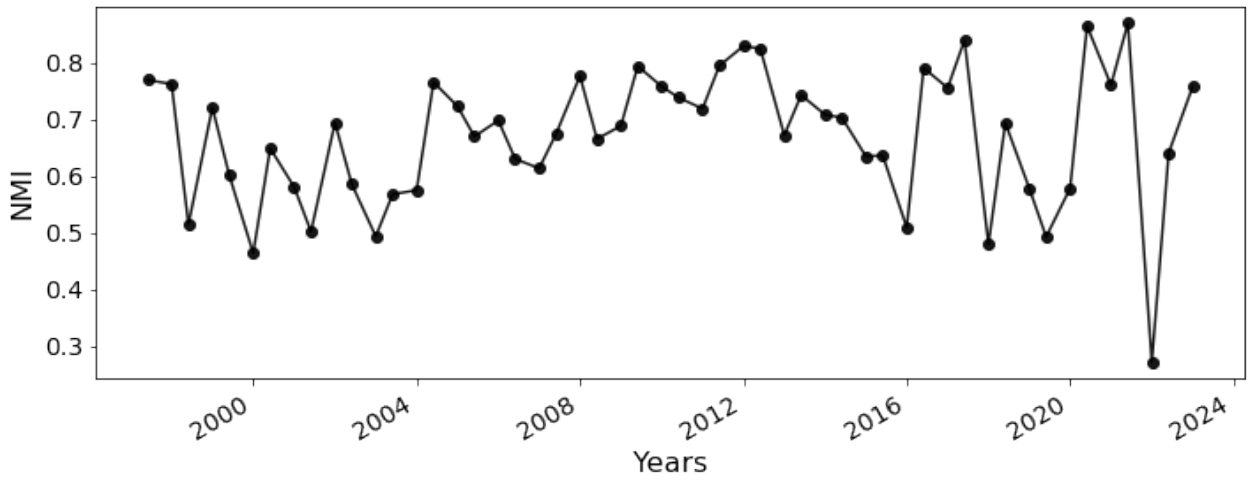
We construct a temporal network by tracking the average intensity of all news between countries (we effective tracking two distinct edges between countries, depending on the ordering of the Source Country and Target Country tuple). We create an edge between countries tracking average news intensity between them, $N_{i,j,t}$. We apply a threshold of at least thirty articles in a given year for a given country source - target relationship to be included within our network in order to focus on meaningful and significant relationships. We choose to take the monthly average across country pairs, rather than treat every piece of news in

the ICEWS dataset as a distinct edge, due to the language biases and over-reporting biases for certain countries and regions within the dataset (Raleigh et al. 2023) – we further justify this through exploring the frequency of reporting news about specific countries in Appendix 7.A.4. The constructed graph is fairly dense, as many pieces of news are written tracking events between countries. We, therefore, set γ^+ and γ^- at one. Our decay parameter is set to 0.9, as we anticipate that older news and events can continue to influence international relations for a significant period of time.

TSL Application Figure 7.13 presents the alluvial diagram for our analysis of the ICEWS international events dataset using the TSL (Boschee et al. 2015). At each time step, we observe the sizes of different clusters - i.e. the number of countries within each cluster. Lighter grey lines represent the flow of countries between clusters from one time period to the next. Figure 7.13a presents the transitions at the bi-annual level. We also track the initial and final clusters of several countries, while the algorithm allows us to pinpoint the relative clustering position of each country at each time step. Figure 7.13b considers the consistency of our extracted clusters across time – we compute the similarity of clusters using the NMI metric from one time step to the next (recomputing clusters every six months). This exercise gives us one way to observe whether international alliances are changing, or whether connections between countries remain stable. Overall, we observe that the period between 1998 and 2004 is marked by relative instability – smaller clusters of countries appear to merge with larger ones. This is consistent with the international turbulence seen within this period, fueled by events such as the September 11, 2001 terrorism attack. After 2004, we observe a long period of relative stability – this time was marked by a shift towards globalization and increased international trade. In 2016, we observe a significant shift as a block of countries in African and the Middle East (such as Pakistan, Lebanon, Sudan, Iraq) appear to merge with a largely Eastern European block (including Russian, Serbia, Belarus). Finally, in 2022



(a) Alluvial Diagram Tracking Country Clusters from International Events



(b) Temporal Cluster Stability Measured through NMI

Figure 7.13: **Alluvial diagram**; at each time step, we observe the sizes of different clusters and the flows between them. In Figure 7.13b we observe the inter-time period stability of clusters (using the NMI to look at cluster consistency) – clusters are recomputed every six months.

we, once again observe a large change as countries consolidated or shifted their alliances with the beginning of the Russian-Ukraine conflict.

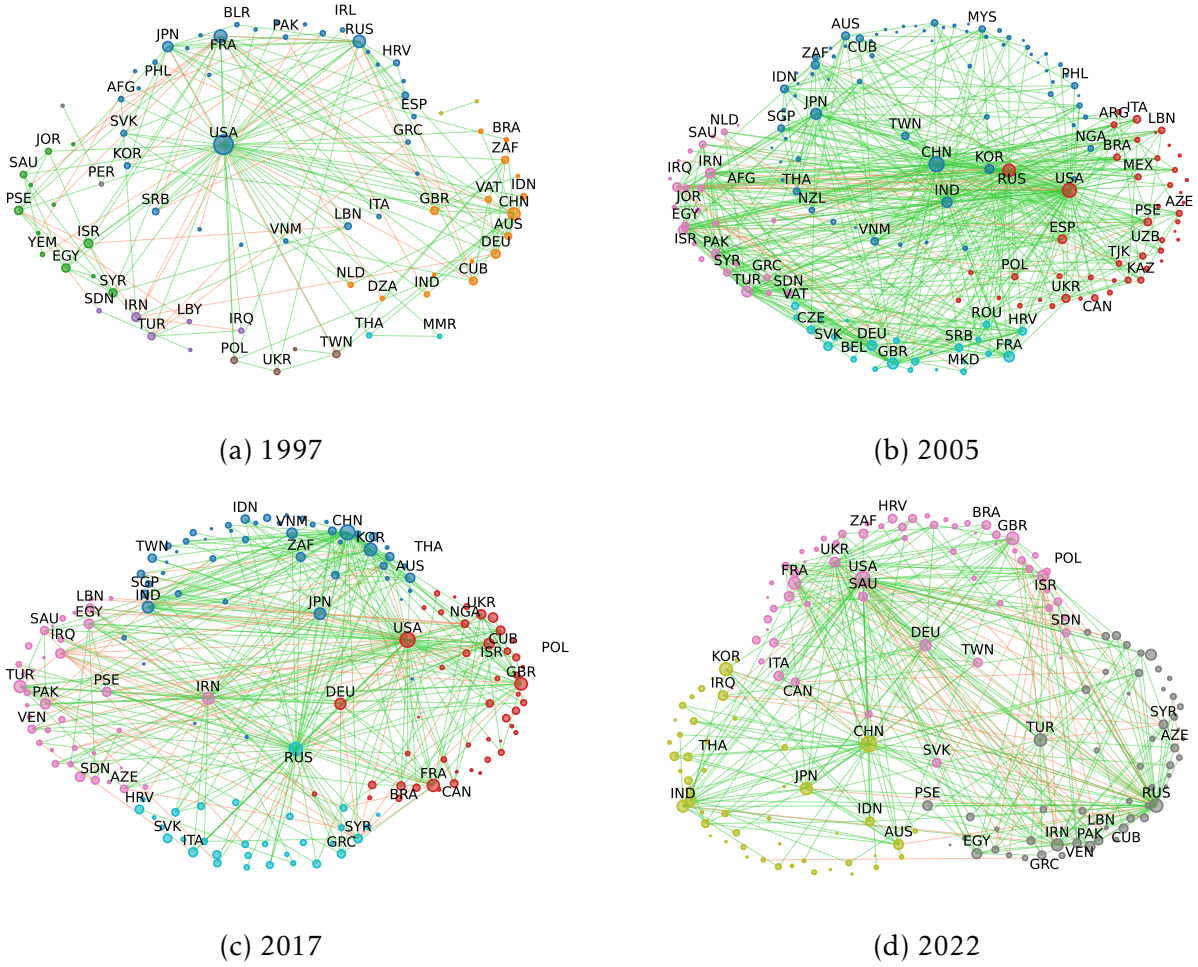


Figure 7.14: **Country Clusters – Network Diagrams**: we present the extracted country clusters across time. Nodes are scaled proportional to their degree and colored according to their temporal clustering allocation. Green edges highlight positive average intensity between countries, while red ones denote negative ones (in Figures 7.14d, 7.14c, 7.14b we only include those with average intensity greater than 0.5 or less than -0.5 due to the increase graph density).

Figure 7.14 shows the results of the clustering exercise. We observe that, generally, geographical proximity is a good indicator of positive relationships. For example, we observe many larger Asian countries clustering together consistently over time, such as China, South Korea and India. Taiwan is a noteworthy exception – it's placement in the network switches to the cluster with the USA and Germany in 2022. Israel is also a noteworthy exception to this rule, since it switches it's cluster from that with it's neighbours to the cluster with the

USA and Germany by 2017.

7.2.3 Conclusion and Further Work

This section presents a signed, temporal extension of the Louvain modularity-maximizing algorithm (Blondel et al. 2008) – the TSL. The key contributions of the section are twofold. We present the flexible algorithm (as well as a robust testing framework) for researchers to use across different domains and for various temporal, signed network clustering settings. We subsequently apply our algorithms to two different empirical applications – to study political and social movements online (DEBAGREEMENT) and to better understand the evolution of international relations through the lens of news events (Boschee et al. 2015). The algorithm provides various tools and output to easily understand community changes over time and to analyse key changes. Given the relative dearth of temporal, signed community detection algorithms, I believe that this chapter fills an important gap in the literature as well as providing opportunities for future research. Beyond the analysis provided here to understand community changes, the richness of current datasets (such as the text data from social media interactions) provides opportunities to use natural language processing methods in combination with the TSL to better understand and improve the interpretability of results. Overall, this chapter is an important step towards taking advantage of the complex, multi-dimensional data which is becoming increasingly available for research.

Appendix

7.A TSL Appendix

7.A.1 Temporal and Signed Network Clustering – Detailed Overview and Extended Literature Review

Signed Clustering – overview In recent years, several methods have been presented to detect communities in (static) signed graphs. These methods are either based on random walks, spectral clustering, generative methods, or social balance theory. Methods based on random walks (Yang et al. 2007, Zhou et al. 2018) generalise the concept of random walks on unsigned graphs to signed graphs by defining a non-zero probability to walk on negative edges. Spectral clustering algorithms (Chen et al. 2018, Cucuringu et al. 2019, 2021, Mercado et al. 2016, 2017, 2019, Zheng & Skillicorn 2015) exploit the eigenvalues of the signed laplacian. In these methods, the number of clusters is fixed a priori. Generative methods (Jiang 2015, Chen, Wang, Yuan & Tang 2014, Yang et al. 2017) propose Stochastic Block and Probabilistic Mixture Models to model non-overlapping and overlapping communities. Methods based on social balance theory (Anchuri & Magdon-Ismail 2012, Traag & Bruggeman 2009) incorporate frustration minimisation and/or modularity maximisation (Newman 2006) to find communities. Recently, people have begun to explore deep-learning embeddings and graph neural networks for community detection (Shen & Chung 2018).

Temporal Clustering – overview Community detection algorithms for temporal graphs are called dynamic community detection (DCD) algorithms in the literature. A recent survey classifies DCD methods into three categories: instant optimal, temporal trade-off, and cross-time methods (Rossetti & Cazabet 2018).

Instant Optimal methods build directly on static community detection algorithms. They consider that communities existing at time t only depend on the current state of the graph at t, so that static algorithms can be applied at each time t. In a second step, communities from different steps are matched by quantifying their similarities. With these methods, it can be difficult to distinguish between the changes due to the evolution of the community structure and the changes due to the instability of the algorithms (static community detection algorithms can find different communities over several iterations on the same graph) (Rosvall & Bergstrom 2010).

Temporal trade-off methods find communities for the initial state of the graph then iteratively find communities at step t using the graph at time t and the communities found in the previous step(s). These approaches are able to cope with instabilities while smoothing the community evolution. Cross-time methods detect communities in a single process by considering all states of the graph at the same time. A single graph, in which each meta-node corresponds to the presence of a node at a given time, is created. Such methods are not able to handle real-time community detection. All above-mentioned methods are designed for unsigned temporal graphs.

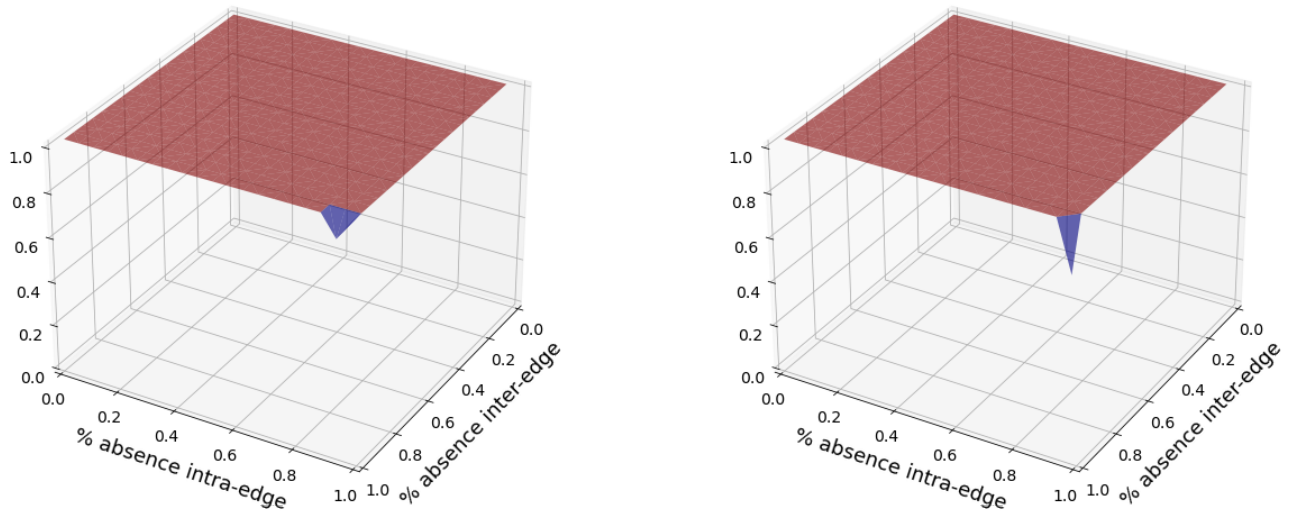
We develop the the TSL – the method has the ability to detect community structures in time-evolving signed graphs, while preserving the continuity of their evolution over the course of time.

7.A.2 Extended Hyperparameter Discussion

Hyperparameter selection: γ

The choice of γ^+ and γ^- is of particular relevance for sparse graphs. We discuss this both empirically and theoretically below.

Consider two nodes u, v in our graph. We consider the (non-normalized) modularity



(a) $\gamma^+ = 0.5, \gamma^- = 0.8$

(b) $\gamma^+ = 0.5, \gamma^- = 1.5$

(c) $\gamma^+ = 1, \gamma^- = 1$

(d) $\gamma^+ = 1.5, \gamma^- = 0.5$

Figure 7.A.1: **NMI for Community Detection in Stochastic Block Model with Edge Sparsity**. The diagrams depict the NMI of the TSL in the static setting as edges are made more sparse within and between communities. The setting has no noise – inter-community edges are always negative, while intra-community edges are always positive. The ability of the alogirthm to identify the true communities from the data generation process is dependent on the resolution parameters γ^+, γ^-.

change for adding the nodes to the same community. If the nodes share a positive link, the modularity change is:

$$1 - \gamma^+ \frac{k_u^+ k_v^+}{2m^+} + \gamma^- \frac{k_u^- k_v^-}{2m^-}. \tag{7.6}$$

A negative link would result in a modularity change of:

$$\gamma^{-}\frac{k_u^{-}k_v^{-}}{2m^{-}} - \gamma^{+}\frac{k_u^{+}k_v^{+}}{2m^{+}} - 1. \tag{7.7}$$

No link would result in a modularity change of:

$$\gamma^{-}\frac{k_u^{-}k_v^{-}}{2m^{-}} - \gamma^{+}\frac{k_u^{+}k_v^{+}}{2m^{+}}. \tag{7.8}$$

The algorithm, therefore, would be incentivized to place nodes in the same communities in the instance when $\gamma^{-} >> \gamma^{+}$.

Additional Synthetic Experiments Indeed, we present the algorithm's performance in our static, signed stochastic block model with no noise from Section 12 with different γ^{+} and γ^{-} parameters (the scenario tests for performance across graphs with varying levels of sparsity). In Figure 7.A.1, we observe that algorithmic performance is indistinguishable in most settings across all γ^{+}, γ^{-} values except for those with few inter and intra community edges. In these cases, we observe better performance when $\gamma^{-} >> \gamma^{+}$.

As is clear from the experiments in Figure 7.A.1 , in the sparse case, domain knowledge can be particularly useful when selecting γ^{+}, γ^{-}. The knowledge that nodes have a greater probability to cluster into larger communities with few observed ties allows us to select a high γ^{-} relative to γ^{-}, such as the combination $\gamma^{+} = 0.5, \gamma^{-} = 0.8$. In our social media application, we consider that the setting is likely similar. The social media forums we observe likely contain several key discussion topics and opinions – while most users may be unlikely to connect with all other users (making the graphs relatively sparse), they do likely belong to a larger overarching community.

Hyperparameter selection: η

Algorithm 2: Time Decay Outline

Run algorithm on all edge that occur before the start time T_0 (as static, signed network). Initialize node cluster allocations C as C_{T_0} and set current time, $T_{current}$ to T_0
for *Each time step when clustering allocations are computed* $T_i \in [T_1, T_2, ...]$ **do**
  for *Each edge that occurs before* T_i **do**
    if $t > T_{i-1}$ **then**
      Compute time for time decay in seconds as $D_{e_t} = T_i - t$
    else
      Compute time for time decay in seconds as $D_{e_t} = T_i - T_{i-1}$
    end
    Update edge weight by multiplying weight of edge by time scaling parameter $= e^{-S*D_{e_t}}$
    Update time t of edge to T_i
    Remove node if the total weight of its edges is below activity threshold
  end
  Compute new node cluster allocations C_{T_i} initializing the computation with allocation from previous time step
  return *Node cluster allocations* C_{T_i}
end

We set the hyperparameter η to decrease the significance of older interactions within our data. The scaling methodology works as follows. The scale parameter S is set to:

$$S = -log(\eta)/(3,600 * 24 * 30 * 3). \tag{7.9}$$

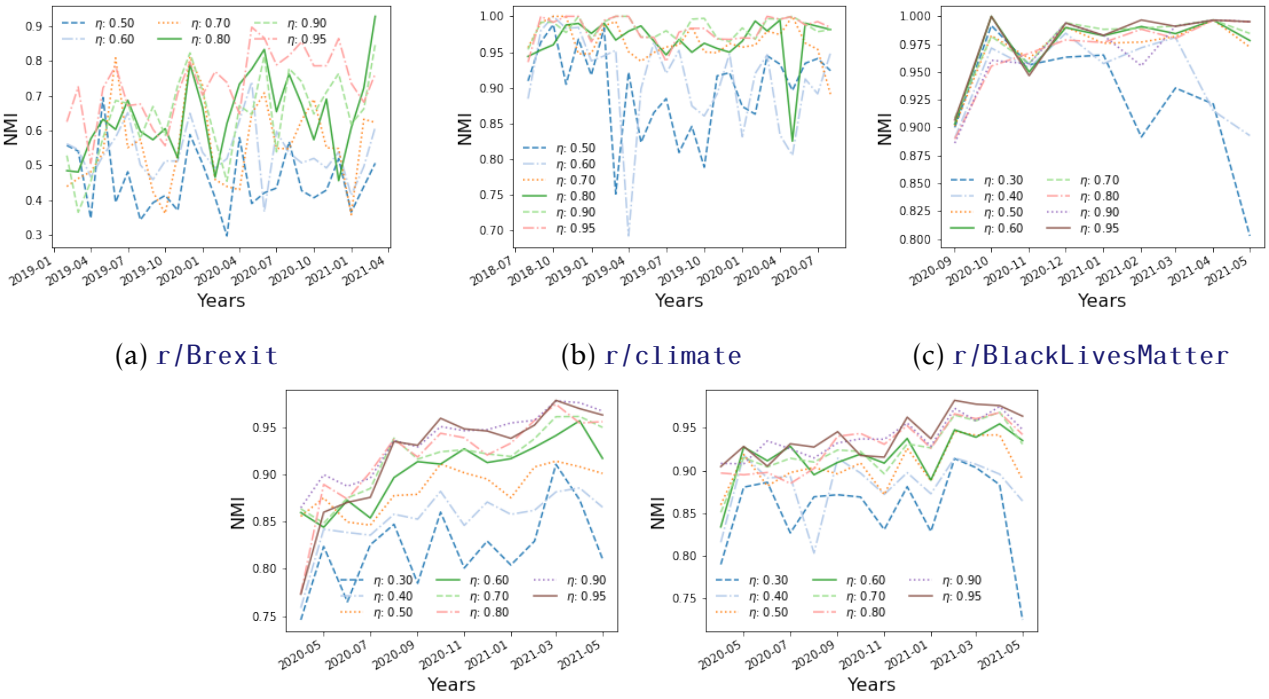
The algorithm processes edges $e_t = (t, u, v, w_{t,u,v}$ according to pseudocode outlined in Algorithm 2.

Let us consider the scaling of an edge that occurs ten seconds before the clustering allocations are computed at T_i: $exp(-10 * -log(0.9)/(3,600 * 24 * 30 * 3)) = 0.9999998645055189$. On the other hand, if the interaction occurred a month before, the weight of the interaction would be: $exp(-3,600 * 24 * 30 * -log(0.9)/(3,600 * 24 * 30 * 3)) = 0.9654893846056298$. At three months before, the weight would be precisely 0.9.

At each time that the algorithm is rerun, the weight for the interaction is multiplied

by the difference between the time at which the interaction occurred and the last time the algorithm was run.

7.A.3 DEBAGREEMENT Application



(a) r/Brexit (b) r/climate (c) r/BlackLivesMatter

(d) r/democrats (e) r/republican

Figure 7.A.2: **DEBAGREEMENT η Selection**: we compare the community structure extracted by the TSL algorithm among active nodes on the forum by comparing the similarity of detected communities at monthly intervals using the NMI metric. We plot the stability of communities for different η parameter values. The goal is to select an η parameter that allows us to detect meaningful community changes, but also is resilient to noise.

Selection of η Hyperparameter We select our time decay parameter η by comparing the NMI among the communities in which active nodes are placed into when updating community allocations using the TSL algorithm at a monthly time horizon – the objective is to select an η parameter that allows us to detect meaningful community changes, but also one that is resilient to noise. Our experiments are presented in Figure 7.A.2.

For r/Brexit and r/climate, we consider slightly higher η parameters (implying that more information will remain from older interactions and information from older ties will

take longer to decay. We select an η of 0.7 for r/Brexit and 0.6 for r/climate. These choices ensure that there are periods of stability over several months – but also allow the algorithm to pick up important changes in community structure (as evidenced by dips in the NMI).

For the forums where we have fewer observations (r/BlackLivesMatter, democrats, r/republicans) and ones that track a particular political discussion – namely for the democrats, r/republicans we observe the period coming up to the presidential elections, while r/BlackLivesMatter considers the development of the movement after George Floyd – we consider a lower η. We want the algorithm to be more adaptable to new information and to capture sudden changes of community structure driven by changes in social movements or political events. For this reason, we select η of 0.4, 0.3, 0.3 for the r/BlackLivesMatter, democrats, r/republicans forums, respectively.

7.A.4 ICEWS Application

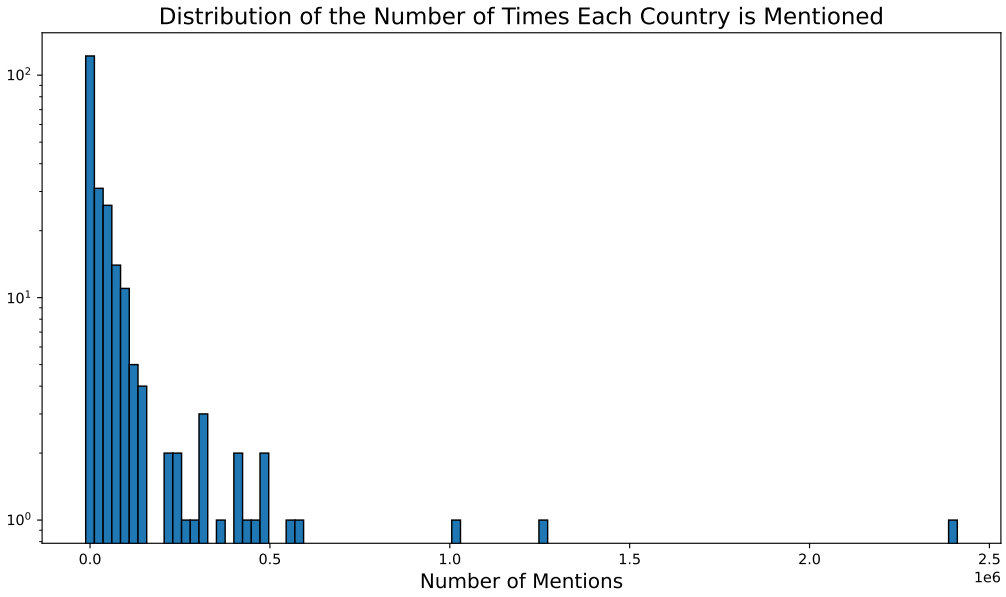


Figure 7.A.3: **Country Mentions in ICEWS**; frequency with which countries are mentioned as either source or target countries for international events, plotted with a log-scaled y-axis.

We justify computing a monthly average inter-country intensity by observing potential

language biases in reporting in ICEWS Raleigh et al. (2023), as well as the much greater magnitude of reporting on some countries. Figure 7.A.3 shows the distribution of the frequency with which a country appears as either the source or the target country in the international events database. We observe that the distribution is heavy tailed – most countries are only mentioned a few times, while some are mentioned hundreds of thousands of times. This work strives to equally represent the importance of all country relations and, therefore, applies the monthly aggregation transformation to the data.

Chapter 8: Conclusion

Social media and internet discussion forums are not, inherently, novel phenomena. In fact, early internet messaging platforms and chat rooms date back to the 1980's and 1990's with the creation of software such as the Internet Relay Chat (intended to help connect people with similar interests), the popularization of AOL's instant messenger AIM and enthusiasm about re-establishing lost connections through platforms such as classmates.com.[1,2] However, it is not until recent years that society has had to reckon with the positive and negative consequences of global interconnectedness and coordination. Cyberbullying, addictive use, the spread of fake news, polarization, election meddling, the spread of extremist views, and privacy abuse are among its documented consequences (Baccarella et al. 2018).

In order to ensure that social media improves welfare and fosters a better society, it is crucial to better understand the dynamics occurring on social media platforms. Specifically, gaining meaningful insights requires: i) the exploration of new, granular datasets documenting underlying trends, ii) the advancement of quantitative and analytical methods targeted towards studying the space, and iii) careful empirical analysis applying novel methods to the collected data. This thesis contributes to each of these efforts, while studying several important areas of society where social media has significantly altered the landscape and where its impact is less understood.

Chapter 3 introduces the WSB datasets and pursues several avenues of analysis. The application of NLP tools and network theoretic techniques uncovers the social and text frameworks which help explain asset correlations. We observe that topics extracted from a LLM applied to text from WSB offer greater context to what drives co-movements in as-

[1] https://www.britannica.com/topic/instant-messaging

[2] https://www.computerworld.com/article/1552589/timeline-the-evolution-of-online-communities.html

sets (Ettinger 2020). For example, pharma companies are linked by the same topics of drug discovery, FDA approvals and the COVID vaccine, in a way that is distinct from the topics that are common to other assets. The analysis also demonstrates the propensity for social investors (active on the WSB forum) to be trend followers – returns in assets experience a run-up for several days leading up to a post, but experience a sharp reversal after the asset is mentioned on WSB.

After establishing a link between WSB and the financial markets our goal is to better understand the underlying dynamics and causal drivers of sentiment formation among social investors, and potential resultant financial market fluctuations. Chapter 4 empirically documents peer effects in expressed sentiments among retail investors on WSB. User sentiments are, on average, 15% more likely to be bullish rather than bearish, if the odds of peers expressing bullish over bearish sentiments double. Our observed group of traders on WSB appear to weight the sentiments of peers more heavily than extrapolation, when forming their expectations for future price movements. Chapter 5 presents several models which document channels through which the behaviours documented among retail investors (peer effects and extrapolation) can destabilize markets – I leverage several modeling approaches, ranging from quantitative finance to complex systems methods. Chapter 5 ties observation from WSB to financial markets in three ways. Shifts in attention to different assets on the forum correspond to changes in retail investor trading patterns. In the presence of 'granular' heavy-tailed attention (viral content), *heterogeneous* investment decisions are not averaged out and impact returns. Finally, the chapter highlights that WSB users exhibit a preference for discussing assets that exhibit bubble-like dynamics.

After the GME short squeeze, the usage of retail investment trading platforms skyrocketed, with *Trading 212* temporarily pausing new account openings in February 2021 due to huge demand, and several other providers struggling to cope with the influx of eager retail

investors. Others have developed new features to allow traders to seamlessly execute on the psychological biases explored within this paper: eToro, for example, now offers a *CopyTrader* feature allowing users to precisely mimic the portfolios of others. Other discussion forums similar to WSB are rising to prominence: for example, the new forum r/StockMarketLeakz was one of the top-growing forums on subreddit in July, 2023. Given the recent trends, it is likely that we have only seen the tip of the iceberg, in terms of the impact that social media and hype investors can have on the market. Despite the thorough analysis of the WallStreetBets forum presented in this thesis, I believe that we have far from a comprehensive understanding of the effects that the influx of social retail traders may have on financial markets and on wealth redistribution more broadly. One particularly disturbing example is the recent explosion in retail trading volumes of zero-day-to-expiry (0DTE) options.[3] Such financial instruments are unlikely to be the investment tools of choice for long-term, sound financial planners, but rather are indicative of gambling – the approach of Bowman et al. (2020) may suggest more appropriate tools for research, rather than traditional financial rational expectations models. All of these trends and changes point to the need for new techniques that incorporate a greater diversity of psychological and cognitive biases, as well as collective dynamics (Garcia et al. 2024), and consider the potential for chaotic dynamics / phase transitions in a population with many, idiosyncratic agents.

As researchers, our modeling abilities are limited by our empirical observations and knowledge of the system we are trying to model. In the recent years, there has been an emphasis on using the controlled experiments to prove causality when studying human behaviour, while underestimating the potential of natural experiments. This trend has led to the neglect of some of the problems with the controlled set up – such as the fact that many experiments are not reproducible, are highly sensitive to the specific environment (with similar experiments generating disparate findings) and potentially fail to translate

[3] https://www.spiderrock.net/an-explosive-combo-zero-day-options-and-retail-traders/

to alternative settings (Almaatouq et al. 2024, Van Bavel et al. 2016). On the other hand, the greater availability of novel, unstructured datasets and powerful computational tools to extract signal have furthered the usefulness of real-world datasets. This thesis strives to demonstrate the usefulness of text data specifically, and to foster greater research using text and social media data for research.

In summary, this thesis is written at a time when new social media data provides opportunity for greater transparency, as well as more realistic modeling of individual investor behaviours and the potential instability caused by their collective dynamics. Unfortunately, a focus on the controlled experimental setting has prevented researchers from unlocking the full potential of this novel data. In a similar vein, traditional, rational-expectations models for investor behaviour may no longer be the appropriate tool for modeling social collective dynamics online, which are increasing in momentum and likely to continue to drive financial market fluctuations.

Chapter 7 broadens the scope of the thesis by introducing a novel social media dataset, DEBAGREEMENT, as well as a new temporal clustering approach for better understanding social movements. These two contributions allow me to to study political and social movements online and to better understand the evolution of international relations through the lens of news events (Boschee et al. 2015). The work demonstrates the growing potential for network theoretic tools to uncover novel insights – especially as our understanding of how different systems can be represented as networks continues to grow. Consider, for example the recent success of deep-learning network models at predicting Drug Target Interactions through representing molecular interactions as networks, or improvements in AI systems from the use of 'knowledge graphs', which record real-world entities or concepts and record relationships between them as a multi-dimensional network (Zhao et al. 2021, Chen, Jia & Xiang 2020, Fensel et al. 2020). A future direction for my research strives to explore

the insights we can gain from applying deep-learning on graphs to financial and economic networks – such as predicting future growth and specialization trajectories for countries.

It is my ambition that social media serves our society – strengthening inter-personal connections, enhancing individual freedoms and promoting widespread access to knowledge. In its current form, social media does not accomplish these goals. A silver lining is the granular datasets available from the usage of online platforms, enabling novel research. The aim of this thesis is twofold: to enhance our understanding of the social-media ecosystem and inform future policy / design decisions, and to advance the mathematical tools and analyses at our disposal to enable more effective use of unstructured (text and network) datasets.

Bibliography

Agrawal, S., Azar, P. D., Lo, A. W. & Singh, T. (2018), 'Momentum, mean-reversion, and social media: Evidence from stocktwits and twitter', *The Journal of Portfolio Management* **44**(7), 85–95.

Al-Andoli, M. N., Tan, S. C., Cheah, W. P. & Tan, S. Y. (2021), 'A review on community detection in large complex networks from conventional to deep learning methods: a call for the use of parallel meta-heuristic algorithms', *IEEE Access* **9**, 96501–96527.

AlDayel, A. & Magdy, W. (2021), 'Stance detection on social media: State of the art and trends', *Information Processing & Management* **58**(4), 102597.

Almaatouq, A., Griffiths, T. L., Suchow, J. W., Whiting, M. E., Evans, J. & Watts, D. J. (2024), 'Beyond playing 20 questions with nature: Integrative experiment design in the social and behavioral sciences', *Behavioral and Brain Sciences* **47**, e33.

An, L., Lou, D. & Shi, D. (2022), 'Wealth redistribution in bubbles and crashes', *Journal of Monetary Economics* **126**, 134–153.

Anchuri, P. & Magdon-Ismail, M. (2012), Communities and balance in signed networks: A spectral approach, *in* '2012 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining', IEEE, pp. 235–242.

Anderson, C. A., Lepper, M. R. & Ross, L. (1980), 'Perseverance of social theories: The role of explanation in the persistence of discredited information.', *Journal of Personality and Social Psychology* **39**(6), 1037.

Angelov, D. (2020), 'Top2vec: Distributed representations of topics', *arXiv preprint arXiv:2008.09470* .

Angrist, J. D. (2014), 'The perils of peer effects', *Labour Economics* **30**, 98–108.

Antweiler, W. & Frank, M. Z. (2004), 'Is all that talk just noise? the information content of internet stock message boards', *The Journal of finance* **59**(3), 1259–1294.

Araci, D. (2019), 'Finbert: Financial sentiment analysis with pre-trained language models', *arXiv preprint arXiv:1908.10063* .

Aral, S., Muchnik, L. & Sundararajan, A. (2009), 'Distinguishing influence-based contagion from homophily-driven diffusion in dynamic networks', *Proceedings of the National Academy of Sciences* **106**(51), 21544–21549.

Avery, C. & Zemsky, P. (1998), 'Multidimensional uncertainty and herd behavior in financial markets', *American Economic Review* pp. 724–748.

Azar, P. D. & Lo, A. W. (2016), 'The wisdom of twitter crowds: Predicting stock market reactions to fomc meetings via twitter feeds', *The Journal of Portfolio Management* **42**(5), 123–134.

Bacaksizlar Turbic, N. G. & Galesic, M. (2023), 'Group threat, political extremity, and collective dynamics in online discussions', *Scientific Reports* **13**(1), 2206.

Baccarella, C. V., Wagner, T. F., Kietzmann, J. H. & McCarthy, I. P. (2018), 'Social media? it's serious! understanding the dark side of social media', *European Management Journal* **36**(4), 431–438.

Bail, C. A., Argyle, L. P., Brown, T. W., Bumpus, J. P., Chen, H., Hunzaker, M. F., Lee, J., Mann, M., Merhout, F. & Volfovsky, A. (2018), 'Exposure to opposing views on social media can increase political polarization', *Proceedings of the National Academy of Sciences* **115**(37), 9216–9221.

Balasubramaniam, V., Campbell, J. Y., Ramadorai, T. & Ranish, B. (2023), 'Who owns what? a factor model for direct stockholding', *The Journal of Finance* **78**(3), 1545–1591.

Baly, R., Mohtarami, M., Glass, J., Màrquez, L., Moschitti, A. & Nakov, P. (2018), 'Integrating stance detection and fact checking in a unified corpus', *arXiv preprint arXiv:1804.08012* .

Banerjee, A., Chandrasekhar, A. G., Duflo, E. & Jackson, M. O. (2013), 'The diffusion of microfinance', *Science* **341**(6144).

Banerjee, A. V. (1993), 'The economics of rumours', *The Review of Economic Studies* **60**(2), 309–327.

Bara'a, A. A., Abbood, A. D., Hasan, A. A., Pizzuti, C., Al-Ani, M., Özdemir, S. & Al-Dabbagh, R. D. (2021), 'A review of heuristics and metaheuristics for community detection in complex networks: Current usage, emerging development and future directions', *Swarm and Evolutionary Computation* **63**, 100885.

Barberis, N., Greenwood, R., Jin, L. & Shleifer, A. (2018), 'Extrapolation and bubbles', *Journal of Financial Economics* **129**(2), 203–227.

Baumgartner, J., Zannettou, S., Keegan, B., Squire, M. & Blackburn, J. (2020), The pushshift reddit dataset, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 14, pp. 830–839.

Becker, J., Brackbill, D. & Centola, D. (2017), 'Network dynamics of social influence in the wisdom of crowds', *Proceedings of the national academy of sciences* **114**(26), E5070–E5076.

Bifulco, R., Fletcher, J. M. & Ross, S. L. (2011), 'The effect of classmate characteristics on post-secondary outcomes: Evidence from the add health', *American Economic Journal: Economic Policy* **3**(1), 25–53.

Blei, D. M., Ng, A. Y. & Jordan, M. I. (2003), 'Latent dirichlet allocation', *Journal of machine Learning research* **3**(Jan), 993–1022.

Blondel, V. D., Guillaume, J.-L., Lambiotte, R. & Lefebvre, E. (2008), 'Fast unfolding of communities in large networks', *Journal of statistical mechanics: theory and experiment* **2008**(10), P10008.

Boehmer, E., Jones, C. M., Zhang, X. & Zhang, X. (2021), 'Tracking retail investor activity', *The Journal of Finance* **76**(5), 2249–2305.

Bond, R. M., Fariss, C. J., Jones, J. J., Kramer, A. D., Marlow, C., Settle, J. E. & Fowler, J. H. (2012), 'A 61-million-person experiment in social influence and political mobilization', *Nature* **489**(7415), 295–298.

Bordalo, P., Gennaioli, N., Kwon, S. Y. & Shleifer, A. (2021), 'Diagnostic bubbles', *Journal of Financial Economics* **141**(3), 1060–1077.

Borge-Holthoefer, J., Magdy, W., Darwish, K. & Weber, I. (2015), Content and network dynamics behind egyptian political polarization on twitter, *in* 'Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work and Social Computing', CSCW '15, Association for Computing Machinery, New York, NY, USA, p. 700–711.
URL: *https://doi.org/10.1145/2675133.2675163*

Boschee, E., Lautenschlager, J., Shellman, S. & Shilliday, A. (2015), 'ICEWS Dictionaries'.
URL: *https://doi.org/10.7910/DVN/28118*

Bouchaud, J.-P. (2013), 'Crises and collective socio-economic phenomena: simple models and challenges', *Journal of Statistical Physics* **151**(3), 567–606.

Bowman, C. E., Brown, R. C., Grindrod, P. & Wardle, H. (2020), 'Online gambling: Hidden markov models for behavioural changes'.

Bradley, D., Hanousek Jr, J., Jame, R. & Xiao, Z. (2024), 'Place your bets? the value of investment research on reddit's wallstreetbets', *The Review of Financial Studies* **37**(5), 1409–1459.

Bramoullé, Y., Djebbari, H. & Fortin, B. (2020), 'Peer effects in networks: A survey', *Annual Review of Economics* **12**, 603–629.

Breza, E. & Chandrasekhar, A. G. (2019), 'Social networks, reputation, and commitment: evidence from a savings monitors experiment', *Econometrica* **87**(1), 175–216.

Brock, W. A. & Durlauf, S. N. (2001), 'Discrete choice with social interactions', *The Review of Economic Studies* **68**(2), 235–260.

Bursztyn, L., Ederer, F., Ferman, B. & Yuchtman, N. (2014), 'Understanding mechanisms underlying peer effects: Evidence from a field experiment on financial decisions.', *Econometrica* **82**(4), 1273.

Buz, T. & de Melo, G. (2021), 'Should you take investment advice from wallstreetbets? a data-driven approach', *arXiv preprint arXiv:2105.02728* .

Cammaerts, B. (2015), 'Social media and activism', *The International Encyclopedia of Digital Communication and Society* pp. 1027–1034.

Cartwright, D. & Harary, F. (1956), 'Structural balance: a generalization of heider's theory.', *Psychological review* **63**(5), 277.

Chen, H., De, P., Hu, Y. J. & Hwang, B.-H. (2014), 'Wisdom of crowds: The value of stock opinions transmitted through social media', *The Review of Financial Studies* **27**(5), 1367–1403.

Chen, H. & Hwang, B.-H. (2022), 'Listening in on investor's thoughts and conversations', *Journal of Financial Economics* **145**(2, Part B), 426–444.

Chen, J., Liu, D., Hao, F. & Wang, H. (2020), 'Community detection in dynamic signed network: an intimacy evolutionary clustering algorithm', *Journal of Ambient Intelligence and Humanized Computing* **11**, 891–900.

Chen, J., Wang, H., Wang, L. & Liu, W. (2016), 'A dynamic evolutionary clustering perspective: Community detection in signed networks by reconstructing neighbor sets', *Physica A: Statistical Mechanics and its Applications* **447**, 482–492.

Chen, S., Khashabi, D., Yin, W., Callison-Burch, C. & Roth, D. (2019), 'Seeing things from a different angle: Discovering diverse perspectives about claims', *arXiv preprint arXiv:1906.03538* .

Chen, X., Jia, S. & Xiang, Y. (2020), 'A review: Knowledge reasoning over knowledge graph', *Expert systems with applications* **141**, 112948.

Chen, Y., Qian, T., Liu, H. & Sun, K. (2018), " bridge" enhanced signed directed network embedding, *in* 'Proceedings of the 27th ACM international conference on information and knowledge management', pp. 773–782.

Chen, Y., Wang, X., Yuan, B. & Tang, B. (2014), 'Overlapping community detection in networks with positive and negative links', *Journal of Statistical Mechanics: Theory and Experiment* **2014**(3), P03021.

Christakis, N. A. & Fowler, J. H. (2013), 'Social contagion theory: examining dynamic social networks and human behavior', *Statistics in medicine* **32**(4), 556–577.

Cignarella, A. T., Lai, M., Bosco, C., Patti, V., Paolo, R. et al. (2020), Sardistance@ evalita2020: Overview of the task on stance detection in italian tweets, *in* 'EVALITA 2020 Seventh Evaluation Campaign of Natural Language Processing and Speech Tools for Italian', Ceur, pp. 1–10.

Colon-Hernandez, P., Havasi, C., Alonso, J., Huggins, M. & Breazeal, C. (2021), 'Combining pre-trained language models and structured knowledge', *arXiv preprint arXiv:2101.12294* .

Cont, R. & Bouchaud, J.-P. (2000), 'Herd behavior and aggregate fluctuations in financial markets', *Macroeconomic Dynamics* **4**(2), 170–196.

Cookson, J. A., Engelberg, J. E. & Mullins, W. (2022), 'Echo Chambers', *The Review of Financial Studies* **36**(2), 450–500.

Cookson, J. A., Mullins, W. & Niessner, M. (2024), 'Social media and finance', *Available at SSRN 4806692* .

Cucuringu, M., Davies, P., Glielmo, A. & Tyagi, H. (2019), Sponge: A generalized eigenproblem for clustering signed networks, *in* 'The 22nd International Conference on Artificial Intelligence and Statistics', PMLR, pp. 1088–1098.

Cucuringu, M., Singh, A. V., Sulem, D. & Tyagi, H. (2021), 'Regularized spectral methods for clustering signed networks', *Journal of Machine Learning Research* **22**(264), 1–79.

Darwish, K., Magdy, W., Rahimi, A., Baldwin, T. & Abokhodair, N. (2018), 'Predicting online islamophopic behavior after# parisattacks', *The Journal of Web Science* **4**.

Darwish, K., Magdy, W. & Zanouda, T. (2017), Improved stance prediction in a user similarity feature space, *in* 'Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017', pp. 145–148.

Darwish, K., Stefanov, P., Aupetit, M. & Nakov, P. (2020), Unsupervised user stance detection on twitter, *in* 'Proceedings of the International AAAI Conference on Web and Social Media', Vol. 14, pp. 141–152.

De Benedictis, L. & Tajoli, L. (2011), 'The world trade network', *The World Economy* **34**(8), 1417–1454.

Dean Pomerleau, D. R. (2017), 'Fake news challenge'.
URL: *http://www.fakenewschallenge.org/*

Defferrard, M., Bresson, X. & Vandergheynst, P. (2016), 'Convolutional neural networks on graphs with fast localized spectral filtering', *arXiv preprint arXiv:1606.09375* .

Derr, T., Ma, Y. & Tang, J. (2018), Signed graph convolutional networks, *in* '2018 IEEE International Conference on Data Mining (ICDM)', IEEE, pp. 929–934.

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018), 'Bert: Pre-training of deep bidirectional transformers for language understanding', *arXiv preprint arXiv:1810.04805* .

Dey, K., Shrivastava, R., Kaushik, S. & Mathur, V. (2017), Assessing the effects of social familiarity and stance similarity in interaction dynamics, *in* 'International Conference on Complex Networks and their Applications', Springer, pp. 843–855.

Duflo, E., Dupas, P. & Kremer, M. (2011), 'Peer effects, teacher incentives, and the impact of tracking: Evidence from a randomized evaluation in Kenya', *American Economic Review* **101**(5), 1739–74.

Epple, D. & Romano, R. E. (2011), Peer effects in education: A survey of the theory and evidence, *in* 'Handbook of social economics', Vol. 1, Elsevier, pp. 1053–1163.

Ettinger, A. (2020), 'What bert is not: Lessons from a new suite of psycholinguistic diagnostics for language models', *Transactions of the Association for Computational Linguistics* **8**, 34–48.

Fama, E. F. & French, K. R. (2004), 'The capital asset pricing model: Theory and evidence', *Journal of Economic Perspectives* **18**(3), 25–46.

Farmer, J. D. (2002), 'Market force, ecology and evolution', *Industrial and Corporate Change* **11**(5), 895–953.

Fensel, D., Şimşek, U., Angele, K., Huaman, E., Kärle, E., Panasiuk, O., Toma, I., Umbrich, J., Wahler, A., Fensel, D. et al. (2020), 'Introduction: what is a knowledge graph?', *Knowledge graphs: Methodology, tools and selected use cases* pp. 1–10.

Fortunato, S. & Hric, D. (2016), 'Community detection in networks: A user guide', *Physics reports* **659**, 1–44.

Gabaix, X. & Koijen, R. S. (2021), In search of the origins of financial fluctuations: The inelastic markets hypothesis, Technical report, National Bureau of Economic Research.

Gabaix, X. & Koijen, R. S. (2023), 'Granular instrumental variables', *Journal of Political Economy (forthcoming)* .

Galaasen, S., Jamilov, R., Juelsrud, R. & Rey, H. (2020), Granular credit risk, Technical report, National Bureau of Economic Research.

Garber, P. M. (1989), 'Tulipmania', *Journal of Political Economy* **97**(3), 535–560.

Garcia, D., Galesic, M. & Olsson, H. (2024), 'The psychology of collectives', *Perspectives on Psychological Science* **19**(2), 316–319.

Gentzkow, M., Kelly, B. & Taddy, M. (2019), 'Text as data', *Journal of Economic Literature* **57**(3), 535–74.

Glaeser, E. L. & Nathanson, C. G. (2017), 'An extrapolative model of house price dynamics', *Journal of Financial Economics* **126**(1), 147–170.

Golub, B. & Jackson, M. O. (2010), 'Naive learning in social networks and the wisdom of crowds', *American Economic Journal: Microeconomics* **2**(1), 112–149.

Greenwood, R., Shleifer, A. & You, Y. (2019), 'Bubbles for fama', *Journal of Financial Economics* **131**(1), 20–43.

Grindrod, P. & Bovet, A. (2022), 'Organization and evolution of the uk far-right network on telegram', *Applied Network Science* **7**.

Grover, A. & Leskovec, J. (2016), node2vec: Scalable feature learning for networks, *in* 'Proceedings of the 22nd ACM SIGKDD international conference on Knowledge discovery and data mining', pp. 855–864.

Hamilton, W. L. (2017), 'Graph representation learning', *Synthesis Lectures on Artificial Intelligence and Machine Learning* **14**(3), 1–159.

Hamilton, W. L., Ying, R. & Leskovec, J. (2017), 'Inductive representation learning on large graphs', *arXiv preprint arXiv:1706.02216* .

Hammond, D. K., Vandergheynst, P. & Gribonval, R. (2011), 'Wavelets on graphs via spectral graph theory', *Applied and Computational Harmonic Analysis* **30**(2), 129–150.

He, P., Liu, X., Gao, J. & Chen, W. (2020), 'Deberta: Decoding-enhanced bert with disentangled attention', *arXiv preprint arXiv:2006.03654* .

Heimer, R. Z. (2016), 'Peer pressure: Social interaction and the disposition effect', *The Review of Financial Studies* **29**(11), 3177–3209.

Hellwig, C. & Veldkamp, L. (2009), 'Knowing what others know: Coordination motives in information acquisition', *The Review of Economic Studies* **76**(1), 223–251.

Hirshleifer, D. (2020), 'Presidential address: Social transmission bias in economics and finance', *The Journal of Finance* **75**(4), 1779–1831.

Holland, P. W., Laskey, K. B. & Leinhardt, S. (1983), 'Stochastic blockmodels: First steps', *Social networks* **5**(2), 109–137.

Hommes, C. (2013), *Behavioral Rationality and Heterogeneous Expectations in Complex Economic Systems*, Cambridge University Press.

Hommes, C. (2021), 'Behavioral and experimental macroeconomics and policy analysis: A complex systems approach', *Journal of Economic Literature* **59**(1), 149–219.

Hu, D., Jones, C. M., Zhang, V. & Zhang, X. (2021), The rise of reddit: How social media affects retail investors and short-sellers' roles in price discovery, Technical report, SSRN.

Hutto, C. & Gilbert, E. (2014), Vader: A parsimonious rule-based model for sentiment analysis of social media text, *in* 'Proceedings of the international AAAI conference on web and social media', Vol. 8, pp. 216–225.

Jiang, A. Q., Sablayrolles, A., Mensch, A., Bamford, C., Chaplot, D. S., Casas, D. d. l., Bressand, F., Lengyel, G., Lample, G., Saulnier, L. et al. (2023), 'Mistral 7b', *arXiv preprint arXiv:2310.06825* .

Jiang, J. Q. (2015), 'Stochastic block model and exploratory analysis in signed networks', *Physical Review E* **91**(6).
URL: *http://dx.doi.org/10.1103/PhysRevE.91.062805*

Jusup, M., Holme, P., Kanazawa, K., Takayasu, M., Romić, I., Wang, Z., Geček, S., Lipić, T., Podobnik, B., Wang, L. et al. (2022), 'Social physics', *Physics Reports* **948**, 1–148.

Keskar, N. S., McCann, B., Varshney, L. R., Xiong, C. & Socher, R. (2019), 'Ctrl: A conditional transformer language model for controllable generation', *arXiv preprint arXiv:1909.05858* .

Kipf, T. N. & Welling, M. (2016), 'Semi-supervised classification with graph convolutional networks', *arXiv preprint arXiv:1609.02907* .

Kirman, A. (1993), 'Ants, rationality, and recruitment', *The Quarterly Journal of Economics* **108**(1), 137–156.

Kivelä, M., Arenas, A., Barthelemy, M., Gleeson, J. P., Moreno, Y. & Porter, M. A. (2014), 'Multilayer networks', *Journal of complex networks* **2**(3), 203–271.

Kubin, E. & Von Sikorski, C. (2021), 'The role of (social) media in political polarization: a systematic review', *Annals of the International Communication Association* **45**(3), 188–206.

Kuchler, T. & Stroebel, J. (2021), 'Social finance', *Annual Review of Financial Economics* **13**, 37–55.

Küçük, D. & Can, F. (2020), 'Stance detection: A survey', *ACM Computing Surveys (CSUR)* **53**(1), 1–37.

Kulkarni, V., Mishra, S. & Haghighi, A. (2021), 'Lmsoc: An approach for socially sensitive pretraining', *arXiv preprint arXiv:2110.10319* .

Lahno, A. M. & Serra-Garcia, M. (2015), 'Peer effects in risk taking: Envy or conformity?', *Journal of Risk and Uncertainty* **50**(1), 73–95.

Lamberson, P., Page, S. E. et al. (2012), 'Tipping points', *Quarterly Journal of Political Science* **7**(2), 175–208.

Lazaridou, A., Kuncoro, A., Gribovskaya, E., Agrawal, D., Liska, A., Terzi, T., Gimenez, M., d'Autume, C. d. M., Ruder, S., Yogatama, D. et al. (2021), 'Pitfalls of static language modelling', *arXiv preprint arXiv:2102.01951* .

Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010*a*), Predicting positive and negative links in online social networks, *in* 'Proceedings of the 19th international conference on World wide web', pp. 641–650.

Leskovec, J., Huttenlocher, D. & Kleinberg, J. (2010*b*), Signed networks in social media, *in* 'Proceedings of the SIGCHI conference on human factors in computing systems', pp. 1361–1370.

Liu, Y., Ott, M., Goyal, N., Du, J., Joshi, M., Chen, D., Levy, O., Lewis, M., Zettlemoyer, L. & Stoyanov, V. (2019), 'Roberta: A robustly optimized bert pretraining approach', *arXiv preprint arXiv:1907.11692* .

Lucchini, L., Aiello, L. M., Alessandretti, L., De Francisci Morales, G., Starnini, M. & Baronchelli, A. (2022), 'From reddit to wall street: The role of committed minorities in financial collective action', *Royal Society Open Science* **9**(4), 211488.

Lux, T. (1998), 'The socio-economic dynamics of speculative markets: interacting agents, chaos, and the fat tails of return distributions', *Journal of Economic Behavior & Organization* **33**(2), 143–165.

MacKinnon, J. G. & White, H. (1985), 'Some heteroskedasticity-consistent covariance matrix estimators with improved finite sample properties', *Journal of Econometrics* **29**(3), 305–325.

Mancini, A., Desiderio, A., Di Clemente, R. & Cimini, G. (2022), 'Self-induced consensus of reddit users to characterise the gamestop short squeeze', *Scientific reports* **12**(1), 13780.

Masuda, N. & Lambiotte, R. (2016), *A guide to temporal networks*, World Scientific.

Mercado, P., Tudisco, F. & Hein, M. (2016), 'Clustering signed networks with the geometric mean of laplacians', *Advances in neural information processing systems* **29**.

Mercado, P., Tudisco, F. & Hein, M. (2017), 'Clustering signed networks with the geometric mean of laplacians'.

Mercado, P., Tudisco, F. & Hein, M. (2019), 'Spectral clustering of signed graphs via matrix power means'.

Meyers, L. A., Newman, M. & Pourbohloul, B. (2006), 'Predicting epidemics on directed contact networks', *Journal of theoretical biology* **240**(3), 400–418.

Monti, F., Boscaini, D., Masci, J., Rodola, E., Svoboda, J. & Bronstein, M. M. (2017), Geometric deep learning on graphs and manifolds using mixture model cnns, *in* 'Proceedings of the IEEE conference on computer vision and pattern recognition', pp. 5115–5124.

Moser, S. C. & Dilling, L. (2007), 'Toward the social tipping point: Creating a climate for change', *Creating a climate for change: Communicating climate change and facilitating social change* pp. 491–516.

Newman, M. (2018), *Networks*, Oxford university press.

Newman, M. E. (2004), 'Analysis of weighted networks', *Physical review E* **70**(5), 056131.

Newman, M. E. (2006), 'Modularity and community structure in networks', *Proceedings of the national academy of sciences* **103**(23), 8577–8582.

Nguyen, L. T., Wu, P., Chan, W., Peng, W. & Zhang, Y. (2012), Predicting collective sentiment dynamics from time-series social media, *in* 'Proceedings of the first international workshop on issues of sentiment discovery and opinion mining', pp. 1–8.

OpenAI (2023), 'Chatgpt: Chatbot based on gpt-3', https://www.openai.com/chatgpt. Accessed: 2024-06-24.

Patacchini, E. & Zenou, Y. (2016), 'Social networks and parental behavior in the intergenerational transmission of religion', *Quantitative Economics* **7**(3), 969–995.

Pearson, N. D., Yang, Z. & Zhang, Q. (2021), 'The chinese warrants bubble: Evidence from brokerage account records', *The Review of Financial Studies* **34**(1), 264–312.

Peters, M. E., Neumann, M., Logan IV, R. L., Schwartz, R., Joshi, V., Singh, S. & Smith, N. A. (2019), 'Knowledge enhanced contextual word representations', *arXiv preprint arXiv:1909.04164* .

Pool, V. K., Stoffman, N. & Yonker, S. E. (2015), 'The people in your neighborhood: Social interactions and mutual fund portfolios', *The Journal of Finance* **70**(6), 2679–2732.

Pougué-Biyong, J. & Semenova, V. (2022), 'From micro to macro: Understanding the social dynamics behind political and economic change', *Complexity in SocialMacroeconomics: Special Issue* .

Pougué-Biyong, J., Semenova, V., Matton, A., Han, R., Kim, A., Lambiotte, R. & Farmer, D. (2021), Debagreement: A comment-reply dataset for (dis) agreement detection in online debates, *in* 'Thirty-fifth Conference on Neural Information Processing Systems Datasets and Benchmarks Track (Round 2)'.

Qiu, X., Sun, T., Xu, Y., Shao, Y., Dai, N. & Huang, X. (2020), 'Pre-trained models for natural language processing: A survey', *Science China Technological Sciences* pp. 1–26.

Rahaman, I. & Hosein, P. (2018), A method for learning representations of signed networks, *in* 'Proceedings of the 14th International Workshop on Mining and Learning with Graphs (MLG)'.

Raleigh, C., Kishi, R. & Linke, A. (2023), 'Political instability patterns are obscured by conflict dataset scope conditions, sources, and coding choices', *Humanities and Social Sciences Communications* **10**(1), 1–17.

Reichardt, J. & Bornholdt, S. (2006), 'Statistical mechanics of community detection', *Physical review E* **74**(1), 016110.

Ribeiro, M. H., Calais, P. H., Almeida, V. A. & Meira Jr, W. (2017), '" everything i disagree with is# fakenews": Correlating political polarization and spread of misinformation', *arXiv preprint arXiv:1706.05924* .

Rosenthal, S. & McKeown, K. (2015), I couldn't agree more: The role of conversational structure in agreement and disagreement detection in online discussions, *in* 'Proceedings of the 16th Annual Meeting of the Special Interest Group on Discourse and Dialogue', pp. 168–177.

Ross, L., Lepper, M. R. & Hubbard, M. (1975), 'Perseverance in self-perception and social perception: biased attributional processes in the debriefing paradigm.', *Journal of Personality and Social Psychology* **32**(5), 880.

Rossetti, G. & Cazabet, R. (2018), 'Community discovery in dynamic networks: A survey', *ACM Comput. Surv.* **51**(2).
URL: *https://doi.org/10.1145/3172867*

Rosvall, M. & Bergstrom, C. T. (2010), 'Mapping change in large networks', *PloS one* **5**(1), e8694.

Sabherwal, S., Sarkar, S. K. & Zhang, Y. (2011), 'Do internet stock message boards influence trading? Evidence from heavily discussed stocks with no fundamental news', *Journal of Business Finance & Accounting* **38**(9-10), 1209–1237.

Sacerdote, B. (2011), Peer effects in education: How might they work, how big are they and how much do we know thus far?, *in* 'Handbook of the Economics of Education', Vol. 3, Elsevier, pp. 249–277.

Samih, Y. & Darwish, K. (2021), A few topical tweets are enough for effective user stance detection, *in* 'Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume', pp. 2637–2646.

Saxena, A., Prasad, M., Gupta, A., Bharill, N., Patel, O. P., Tiwari, A., Er, M. J., Ding, W. & Lin, C.-T. (2017), 'A review of clustering techniques and developments', *Neurocomputing* **267**, 664–681.

Scholl, M. P., Calinescu, A. & Farmer, J. D. (2021), 'How market ecology explains market malfunction', *Proceedings of the National Academy of Sciences* **118**(26), e2015574118.

Schou, P. K., Bucher, E., Waldkirch, M. & Grünwald, E. (2022), We did start the fire: R/wallstreetbets,'flash movements' and the gamestop short-squeeze, *in* 'Academy of Management Proceedings', Vol. 2022, Academy of Management Briarcliff Manor, NY 10510, p. 14028.

Seabold, S. & Perktold, J. (2010), statsmodels: Econometric and statistical modeling with python, *in* '9th Python in Science Conference'.

Semenova, V., Gorduza, D., Wildi, W., Dong, X. & Zohren, S. (2024), 'Wisdom of the crowds or ignorance of the masses? a data-driven guide to wallstreetbets.', *Journal of Portfolio Management* **50**(4).

Semenova, V. & Winkler, J. (2021), 'Social contagion and asset prices: Reddit's self-organised bull runs', *arXiv preprint arXiv:2104.01847* .

Shen, X. & Chung, F.-L. (2018), 'Deep network embedding for graph representation learning in signed networks', *IEEE transactions on cybernetics* **50**(4), 1556–1568.

Shiller, R. J. (1984), Stock prices and social dynamics, Technical Report 2, The Brookings Institution.

Shiller, R. J. (2005), *Irrational Exuberance: (Second Edition)*, Princeton University Press.

Shiller, R. J. (2017), 'Narrative economics', *American Economic Review* **107**(4), 967–1004.

Strehl, A. & Ghosh, J. (2002), 'Cluster ensembles—a knowledge reuse framework for combining multiple partitions', *Journal of machine learning research* **3**(Dec), 583–617.

Tan, C., Niculae, V., Danescu-Niculescu-Mizil, C. & Lee, L. (2016), Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions, *in* 'Proceedings of the 25th international conference on world wide web', pp. 613–624.

Tang, J., Chang, Y., Aggarwal, C. & Liu, H. (2016), 'A survey of signed network mining in social media', *ACM Computing Surveys (CSUR)* **49**(3), 1–37.

Taulé, M., Pardo, F. M. R., Martí, M. A. & Rosso, P. (2018), Overview of the task on multimodal stance detection in tweets on catalan# 1oct referendum., *in* 'IberEval@ SEPLN', pp. 149–166.

Traag, V. A. & Bruggeman, J. (2009), 'Community detection in networks with positive and negative links', *Physical Review E* **80**(3), 036115.

Traag, V. A., Waltman, L. & Van Eck, N. J. (2019), 'From louvain to leiden: guaranteeing well-connected communities', *Scientific reports* **9**(1), 5233.

Traag, V., Doreian, P. & Mrvar, A. (2019), 'Partitioning signed networks', *Advances in network clustering and blockmodeling* pp. 225–249.

Tucker, J. A., Guess, A., Barberá, P., Vaccari, C., Siegel, A., Sanovich, S., Stukal, D. & Nyhan, B. (2018), 'Social media, political polarization, and political disinformation: A review of the scientific literature', *Political polarization, and political disinformation: a review of the scientific literature (March 19, 2018)* .

Van Bavel, J. J., Mende-Siedlecki, P., Brady, W. J. & Reinero, D. A. (2016), 'Contextual sensitivity in scientific reproducibility', *Proceedings of the National Academy of Sciences* **113**(23), 6454–6459.

Veldkamp, L. L. (2006), 'Media frenzies in markets for financial information', *American Economic Review* **96**(3), 577–601.

Veličković, P., Cucurull, G., Casanova, A., Romero, A., Liò, P. & Bengio, Y. (2018), 'Graph Attention Networks', *International Conference on Learning Representations* .
URL: *https://openreview.net/forum?id=rJXMpikCZ*

Voitalov, I., van der Hoorn, P., van der Hofstad, R. & Krioukov, D. (2019), 'Scale-free networks well done', *Physical Review Research* **1**(3), 033034.

Walker, M., Anand, P., Abbott, R. & Grant, R. (2012), Stance classification using dialogic properties of persuasion, *in* 'Proceedings of the 2012 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies', Association for Computational Linguistics, Montréal, Canada, pp. 592–596.
URL: *https://aclanthology.org/N12-1072*

Wan, X., Yang, J., Marinov, S., Calliess, J.-P., Zohren, S. & Dong, X. (2021), 'Sentiment correlation in financial news networks and associated market movements', *Scientific reports* **11**(1), 3062.

Wang, C. & Luo, B. (2021), Predicting $ gme stock price movement using sentiment from reddit r/wallstreetbets, *in* 'Proceedings of the Third Workshop on Financial Technology and Natural Language Processing', pp. 22–30.

Wang, Q., Mao, Z., Wang, B. & Guo, L. (2017), 'Knowledge graph embedding: A survey of approaches and applications', *IEEE Transactions on Knowledge and Data Engineering* **29**(12), 2724–2743.

Wang, S., Tang, J., Aggarwal, C., Chang, Y. & Liu, H. (2017), Signed network embedding in social media, *in* 'Proceedings of the 2017 SIAM international conference on data mining', SIAM, pp. 327–335.

Wang, X., Song, J., Lu, K. & Wang, X. (2017), 'Community detection in attributed networks based on heterogeneous vertex interactions', *Applied Intelligence* **47**, 1270–1281.

Weidmann, B. & Deming, D. J. (2021), 'Team players: How social skills improve team performance', *Econometrica* **89**(6), 2637–2657.

Welch, I. (2022), 'The wisdom of the robinhood crowd', *The Journal of Finance* **77**(3), 1489–1527.

Wojcieszak, M. (2011), 'Deliberation and attitude polarization', *Journal of Communication* **61**(4), 596–617.

Wu, G.-C. & Baleanu, D. (2015), 'Jacobian matrix algorithm for lyapunov exponents of the discrete fractional maps', *Communications in Nonlinear Science and Numerical Simulation* **22**(1-3), 95–100.

Wu, L., Chen, Y., Shen, K., Guo, X., Gao, H., Li, S., Pei, J. & Long, B. (2021), 'Graph neural networks for natural language processing: A survey', *arXiv preprint arXiv:2106.06090* .

Yang, B., Cheung, W. & Liu, J. (2007), 'Community mining from signed social networks', *IEEE transactions on knowledge and data engineering* **19**(10), 1333–1348.

Yang, B., Liu, X., Li, Y. & Zhao, X. (2017), 'Stochastic blockmodeling and variational bayes learning for signed network analysis', *IEEE Transactions on Knowledge and Data Engineering* **29**(9), 2026–2039.

Yang, Z., Dai, Z., Yang, Y., Carbonell, J., Salakhutdinov, R. R. & Le, Q. V. (2019), 'Xlnet: Generalized autoregressive pretraining for language understanding', *Advances in neural information processing systems* **32**.

Zenou, Y. (2016), 'Key players', *The Oxford Handbook of the Economics of Networks* **11**.

Zhang, J., Tan, L. & Tao, X. (2019), 'On relational learning and discovery in social networks: a survey', *International Journal of Machine Learning and Cybernetics* **10**(8), 2085–2102.

Zhao, T., Hu, Y., Valsdottir, L. R., Zang, T. & Peng, J. (2021), 'Identifying drug–target interactions based on graph convolutional network and deep neural network', *Briefings in bioinformatics* **22**(2), 2141–2150.

Zheng, Q. & Skillicorn, D. B. (2015), Spectral embedding of signed networks, *in* 'Proceedings of the 2015 SIAM international conference on data mining', SIAM, pp. 55–63.

Zhou, J., Li, L., Zeng, A., Fan, Y. & Di, Z. (2018), 'Random walk on signed networks', *Physica A: Statistical Mechanics and its Applications* **508**, 558–566.

Zhu, J., Cui, Y., Liu, Y., Sun, H., Li, X., Pelger, M., Yang, T., Zhang, L., Zhang, R. & Zhao, H. (2021), Textgnn: Improving text encoder via graph neural network in sponsored search, *in* 'Proceedings of the Web Conference 2021', pp. 2848–2857.